

UNIVERSIDAD TÉCNICA FEDERICO SANTA
MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO – CHILE



“Diseño conceptual de un Observatorio Virtual
Astronómico para ALMA”

WALTER RODRIGO FARIÑA PÉREZ

MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO
CIVIL INFORMÁTICO.

PROFESOR GUÍA: MAURICIO SOLAR
PROFESOR CORREFERENTE: MARCELO MENDOZA

NOVIEMBRE - 2013

Dedicado a mis padres, a mi hermana, a la Quidora, a mis amigos, a mis profesores y a mis compañeros, por hacer de esta etapa un muy buen recuerdo.

Resumen

Este estudio analiza las necesidades de los astrónomos chilenos respecto a ALMA, definiendo las bases del diseño que posteriormente será el Observatorio Virtual Chileno. Se hace un estudio de necesidades, obtienen requerimientos, casos de uso y finalmente un modelo de datos, este último acorde a los estándares de IVOA, organización internacional de observatorios virtuales. Así, la comunidad astronómica chilena tendrá acceso global y homogeneizado a los datos astronómicos, posteriormente junto a herramientas para explorarlos y manipularlos.

Palabras Clave: Observatorio Virtual Chileno, IVOA, radioastronomía, ALMA, modelo de datos.

Abstract

This study analyzes the Chilean astronomer's needs, respect to ALMA data, defining the basis of the design of the future Chilean Virtual Observatory. This work starts with a study of the astronomer's needs, then gathers the requirements, use cases and finally, the data model, the last one compliant to IVOA standards, the International Virtual Observatory Alliance. So the Chilean astronomical community has global and uniform access to astronomical data, after this, with tools for exploring and manipulate them.

Keywords: Chilean Virtual Observatory, IVOA, radioastronomy, ALMA, data model.

Glosario

ALMA: Atacama Large Millimeter/Submillimeter Array, radiotelescopio ubicado en Chile, consta de 66 antenas.

ARC: ALMA Regional Center, Centro que distribuye datos producidos por ALMA, existen 3 alrededor del mundo, ninguno en Chile.

ASA: ALMA Science Archive, se encarga de dar acceso a los datos de ALMA con un enfoque científico, hay uno en cada ARC.

LSST: Large Synoptic Survey Telescope, Telescopio que para el 2022 estará completamente operativo en Chile y que producirá catálogos de 10 PB en sus 10 años de observaciones.

PB: PetaByte, unidad de almacenamiento de información digital, es equivalente a 10^{15} bytes.

OV: Observatorio Virtual.

ChiVO: Chilean Virtual Observatory.

IVOA: International Virtual Observatory Alliance.

Big data: Término utilizado sobre una colección de conjuntos de datos que son tan grandes y complejos que se vuelve difícil de procesar con herramientas típicas de gestión de bases de datos.

XML: Extensible Markup Language es un lenguaje de marcado que define un conjunto de reglas para codificar documentos en un formato que sea legible por humanos y máquinas.

ICRS: Sistema estándar para referencia celestial utilizado por la IAU.

J2000: Instante en el tiempo utilizado como referencia para sistema de coordenadas.

Sesame: Servicio resolutor de nombres astronómicos, al entregarle un nombre de un objeto retorna las coordenadas de este.

SIMBAD, ADS, NED: Servicios de información bibliográfica que permiten buscar bibliografía respecto a parámetros astronómicos como coordenadas o características físicas.

FITS: Tipo de archivo más utilizado en astronomía para almacenar y transferir datos de observaciones astronómicas.

VOTable: Estructura de almacenamiento de metadatos y datos astronómicos establecido por el IVOA.

ASDM: ALMA Science Data Model, Estructura de archivos que describen una observación realizada por ALMA.

Execution Block: Bloque de ejecución, es la unidad lógica más grande dentro de ALMA que contiene datos de una observación.

Spectral Window: Rango de frecuencias observadas por ALMA.

CASA: Common Astronomy Software Applications, paquete de aplicaciones de Software para la astronomía que está siendo desarrollado con el objetivo principal de proveer soporte al pos procesamiento de datos generados por la próxima generación de radiotelescopios como ALMA.

Índice general

Resumen	ii
Abstract	ii
Glosario	iii
Índice general	v
Índice de figuras	ix
Índice de tablas	ix
Capítulo 1: Introducción	1
1.1 Definición del problema	4
1.1.1. Descripción del problema	4
1.1.2. Alcance	6
1.1.3. Metodología	7
Capítulo 2: Estado del Arte	9
2.1. La Alianza Internacional de Observatorios Virtuales (IVOA)	9
2.1.1 Sección superior arquitectura IVOA	12
2.1.2. Sección intermedia arquitectura IVOA	13
2.1.3. Sección inferior arquitectura IVOA	15
2.2 Estándares de IVOA	16
2.2.1 <i>Observation Core</i>	16
2.2.1.1. Principales conceptos del modelo de datos <i>Observation Core</i>	17
2.2.1.2. Implementación del modelo de datos <i>Observation Core</i> en un servicio de acceso ...	18
2.2.2. Modelo de datos de caracterización (CharDM)	18
2.2.2.1. Principales conceptos del modelo de datos <i>Characterisation</i>	18
2.2.2.2. Ejes y sus atributos	19
2.2.2.3. Cobertura (<i>Coverage</i>)	20
2.2.2.4. Resolución (<i>Resolution</i>) y Precisión de Muestreo	21
2.2.2.5. El Modelo	21
2.2.2.6. Ejes en detalle	21
2.2.2.7. Errores en <i>Characterisation</i> (<i>Accuracy</i>)	22
2.2.2.8. Implementando el modelo usando elementos STC	23
2.2.3. TAP	23
2.2.3.1. Implementaciones	26
2.3. Observatorios Virtuales y el mundo	27
2.4. Estado del Arte Nacional	29
2.4.1. ¿Qué necesita Chile para un OV?	29
2.4.2. Conexiones de redes académicas	29
2.4.3. <i>Grid</i>	30

2.4.4. Creación de <i>Software</i> para OV	31
2.4.5. Software para el Uso de la Grid.....	31
2.4.6. Capital Humano	31
2.4.7. Incorporación al IVOA	31
2.4.8. <i>Roadmap</i>	32
2.4.9. <i>Chilean Virtual Observatory</i>	34
2.5. ALMA.....	35
2.6. Servicios	36
2.6.1. Resolvedores de Nombres	36
2.6.2. Acceso a bases de datos espectroscópicas	38
2.7. Tipos de Archivos.....	39
2.7.1. FITS	39
2.7.2. VOTable.....	39
2.7.3. MS	40
2.7.4. ASDM	40
Capítulo 3: Diseño	43
3.1. Necesidades y Requerimientos	43
3.1.1. Necesidades.....	43
3.1.2. Requerimientos	43
3.2. Casos de Uso.....	47
3.2.1. Actores.....	47
3.2.2. Casos de uso	47
3.3. Diseño de modelo de datos	54
3.3.1. Observación.....	56
3.3.2. Modelo de datos Characterisation de IVOA.....	59
3.3.2.1. Eje Espacial	61
3.3.2.2. Eje Temporal	63
3.3.2.3. Eje Espectral	65
3.3.2.4. Eje Observable	67
3.4 Arquitectura.....	68
3.4.1. Capa de clientes	68
3.4.2. Capa de Aplicaciones	69
3.4.3. Capa de Datos	70
Capítulo 4: Conclusiones	71
4.1. Análisis de los resultados obtenidos	71
4.2. Futuros trabajos	72
Bibliografía	73
Anexo A: Estándares de IVOA	76
A.1. Estándares de Protocolos de obtención y acceso a datos.....	76
A.1.1. SIAP (<i>Simple Image Access Protocol</i>)	76
A.1.2. SCS (<i>Simple Cone Search</i>).....	76
A.1.3. SSAP (<i>Simple Spectra Access Protocol</i>).....	76
A.1.4. SLAP (<i>Simple Line Access Protocol</i>)	77
A.1.5. TAP (<i>Table Access Protocol</i>)	77
A.1.6. SEAP (<i>Simple Event Access Protocol</i>)	77
A.1.7. SimDAL (<i>Simulations Data Access Layer</i>)	77
A.1.8. FAP (<i>Footprint Access Protocol</i>).....	77

A.2. Núcleo del OV: Estándares de Modelos de Datos	78
A.2.1. STC (<i>Space Time Coordinate metadata</i>)	78
A.2.2. <i>Units</i>	78
A.2.3. <i>Utypes</i>	78
A.2.4. <i>SpectrumDM</i>	78
A.2.5. SSLDM (<i>Simple Spectral Line DM</i>)	78
A.2.6. CharDM (<i>Characterization Data Model</i>)	78
A.2.7. ObsCoreDM (<i>Observation Core Data Model</i>)	79
A.2.8. ObsProvDM (<i>Observation and Provenance DM</i>)	79
A.2.9. PhotDM (<i>Photometry DM</i>)	79
A.2.10. SimDM (<i>Simulations Data Model</i>)	79
A.2.11. VOEvent	79
A.3. Núcleo del VO: Estándares de Lenguajes de Consulta	80
A.3.1. ADQL (<i>Astronomical Data Query Language</i>)	80
A.3.2. PQL (<i>Parametrized Query Language</i>)	80
Anexo B: Descripción detallada de columnas del <i>Observation Core</i>.	81
B.1. <i>Data Product Type</i> _(dataproduct_type)_	81
B.2. <i>Calibration Level</i> _(calib_level)_	82
B.3. <i>Collection Name</i> _(obs_collection)_	82
B.4. <i>Observation Identifier</i> _(obs_id)_	83
B.5. <i>Publisher Dataset Identifier</i> _(obs_publisher_did)_	83
B.6. <i>Access URL</i> _(Access_url)_	84
B.7. <i>Access Format</i> _(access_format)_	84
B.8. <i>Estimated Download Size</i> _(access_estsize)_	85
B.9. <i>Target Name</i> _(target_name)_	85
B.10. <i>Cental Coordinates</i> _(s_ra,s_dec)_	85
B.11. <i>Spatial Exent</i> _(s_fov)_	85
B.12. <i>Spatial Coverage</i> _(s_region)_	86
B.13. <i>Spatial Resolution</i> _(s_resolution)_	86
B.14. <i>Time Bounds</i> _(t_min, t_max)_	86
B.15. <i>Exposure Time</i> _(t_exptime)_	86
B.16. <i>Time Resolution</i> _(t_resolution)_	87
B.17. <i>Spectral Bounds</i> _(em_min, em_max)_	87
B.18. <i>Spectral Resolving Power</i> _(em_res_power)_	87
B.19. <i>Observable Axis Description</i> _(o_ucd)_	87
B.20. Columnas adicionales	87
Anexo C: Observatorios Virtuales alrededor del mundo.	88
C.1. Argentina Virtual Observatory	88
C.2. Armenian Virtual Observatory	88
C.3. AstroGrid	89
C.4. Australian Virtual Observatory	89
C.5. Brazilian Virtual Observatory	90
C.6. Canadian Virtual Observatory	91
C.7. Chinese Virtual Observatory	91
C.8. European Space Agency	92
C.9. European Virtual Observatory	93
C.10. German Astrophysical Virtual Observatory	93
C.11. Hungarian Virtual Observatory	94

C.12. Italian Virtual Observatory.....	95
C.13. Japanese Virtual Observatory.....	95
C.14. Observatoire Virtual France.....	95
C.15. Russian Virtual Observatory.....	95
C.16. Spanish Virtual Observatory.....	96
C.17. US Virtual Astronomical Observatory.....	97
C.18. Virtual Observatory India	98

Índice de figuras

2.1. Arquitectura de IVOA.....	12
2.2. Gráfico de observatorios miembros de IVOA.....	28
2.3. Cobertura del espectro electromagnético por distintos instrumentos y OV...28	
2.4. Jerarquía de los “objetos observados” usados en el ASDM.....	41
3.1. UML de Casos de Uso para el ChiVO.	48
3.2. Clases principales del modelo de datos de ChiVO.....	55
3.3. Modelo de Datos de alto nivel del estándar <i>Characterisation</i> para ALMA.....	60
3.4. Eje espacial de la clase de caracterización.....	62
3.5. Eje temporal de la clase de caracterización.....	64
3.6. Eje espectral de la clase de caracterización.....	65
3.7. Eje observable de la clase de caracterización.....	67
3.8. Arquitectura de ChiVO.....	69

Índice de tablas

2.1. Tabla de Campos del <i>Observation Core</i>	17
2.2. Miembros de IVOA.....	27
2.3 Estructura de un VOTable.....	40
3.1. Casos de uso transversales a los requerimientos planteados evaluados por necesidad y prioridad.....	49
3.2. Lista de casos de uso asociados a cada requerimiento evaluados por necesidad y prioridad.....	51
3.3. Lista de campos más relevantes asociados a los requerimientos y ubicación de éstos dentro del ASDM o servicios astronómicos.....	52
3.4. Lista de campos asociados a los requerimientos y ubicación de éstos dentro del ASDM o servicios astronómicos.....	53
3.5. Columnas obligatorias del estándar de modelo de datos <i>Observation Core</i> de IVOA para ser compatible con el ecosistema OV.....	57
3.6. Campos del estándar <i>ObsCoreDM</i> y origen en el caso de ASDM y FITS.....	58

Capítulo 1: Introducción

La astronomía según la Real Academia Española es una “ciencia que trata de cuanto se refiere a los astros, y principalmente a las leyes de sus movimientos” [1], sus primeros indicios se remontan a Mesopotamia, gracias al catálogo de estrellas Babilónico que data aproximadamente del 1200 A.C. Su objetivo principal es el estudio de objetos astronómicos (tales como estrellas, planetas, galaxias) a través de la física (por ejemplo, estudio fenómenos físicos), química (por ejemplo, estudio de moléculas) y evolución (como varían a lo largo del tiempo) así como los fenómenos que suceden afuera de la atmósfera terrestre. Se diferencia de la cosmología cuyo objetivo es entender las leyes generales, del origen y de la evolución del universo [2].

Actualmente, no sólo se estudian los objetos visibles al ojo humano, sino que prácticamente todo el espectro electromagnético incluyendo fuentes emisoras de ondas de radio, radiación ultravioleta, radiación infrarroja y de Rayos X y gamma. Para ello, se utilizan distintas herramientas como telescopios y radiotelescopios que, hoy en día, recopilan información y la almacenan en medios digitales.

El estudio de estos objetos generalmente es realizado por un astrónomo a través de software especializado para ello, donde se analizan los datos obteniendo así la información relevante para comprobar o refutar una o varias hipótesis.

Debido a las tecnologías actuales, la astronomía está sometida a problemas de almacenamiento, transmisión y procesamiento de la gran cantidad de datos digitales que se generan con las observaciones. Actualmente, hay 1 PB de datos públicos y se generan cerca de 0.5PB por año, en los próximos años serán aún más con los proyectos como el LSST (*Large Synoptic Survey Telescope*), ALMA (*Atacama Large Millimeter Array*), y el SKA (*Square Kilometer Array*). Esto requiere mayor cantidad de dispositivos de almacenamiento, lo cual es costoso debido a su gran

volumen de datos, para el 2020 se esperan tener 60PB de datos para acceso público [3].

No sólo el almacenamiento de estos datos es una preocupación, sino también el acceso a esta gran cantidad de datos que se pone a disposición de la comunidad a través de servidores públicos, generalmente en formatos diferentes y distribuidos a y por distintas instituciones. Lo anterior implica que la calidad de los datos, los metadatos, interfaces, y accesibilidad son distintas, debido a que cada proyecto astronómico maneja sus propios datos, los pone a disposición en una base de datos propia, y aún peor, los formatos de los datos en astronomía son dependientes de la instrumentación [4]. Todo esto obliga a los astrónomos a adaptarse a cada fuente de datos de manera independiente, haciéndoles más difícil su trabajo, lo que evidencia la necesidad de crear una arquitectura que los unifique.

La solución encontrada al problema planteado es la creación de Observatorios Virtuales (OV), que son un esfuerzo internacional para ofrecer una integración digital de los datos, herramientas y servicios astronómicos a la comunidad global [4], facilitando así la investigación de los astrónomos de todo el mundo.

Un concepto esencial de los OV es que al proveer un acceso global y homogeneizado a los datos, junto a herramientas para explorarlos y manipularlos, la necesidad por nuevas observaciones no estará directamente relacionada con el aumento de la producción de información científica, esto es debido a que se tendrá acceso a más datos con una mejor capacidad de análisis.

Otra ventaja es que todos los datos recolectados, incluso los datos privados, pueden ser accedidos a través del OV, enriqueciendo a la comunidad internacional. Además, los grandes sondeos realizados por algunos observatorios superarían a las observaciones individuales y puntuales, ofreciendo una coherencia agregada a la estructura del OV. Debido a todo lo anterior, los OV no son una empresa dirigida por

un instituto o incluso un país, es una propuesta de comunidad con el objetivo de democratizar la información que se expandirá a otras áreas científicas como la meteorología, geofísica, y ciencias espaciales, permitiendo nuevas interacciones e intercambios de métodos y tecnologías.

Así, hoy los OV representan a la comunidad astronómica lo que significó el Internet al mundo académico en los ochenta. Es claro hoy que la ciencia, especialmente en los países en vías de desarrollo, podría ser completamente distinta sin el internet de la misma manera que en el futuro se dirá que la astronomía no será lo mismo sin los OV.

Gran parte de la comunidad astronómica está de acuerdo en que el concepto de Observatorio Virtual es la solución para enfrentar los desafíos científicos y tecnológicos impuestos por los conjuntos de datos gigantes y complejos generados por los observatorios astronómicos. Pero los requerimientos de *hardware* es sólo una pequeña parte de los problemas que surgen cuando se trabaja con conjuntos de datos *big data* [7]. Procesar toma grandes cantidades de tiempo, pero una vez terminada esta fase es de gran importancia que las consultas y obtenciones de los datos sea de manera eficiente. Esto requiere inversiones no sólo en *software* de bases de datos, sino también de profesionales en astronomía y computación capaces de trabajar en conjunto para así diseñar e implementar modelos de datos eficientes y científicamente orientados. Esta información, una vez estructurada en tales bases de datos, necesita ser extraída eficientemente, lo que implica conexiones de red de alta velocidad [4].

En los últimos años ha habido un esfuerzo de toda la comunidad astronómica de publicar cada vez más datos en el ecosistema de OV, así es como esta memoria plantea partir con el observatorio radioastronómico *Atacama Large Millimeter/submillimeter Array* (ALMA) ya que actualmente tiene muy poca presencia en los OV y pocas herramientas de análisis en el ecosistema.

La problemática anterior aplica a la comunidad de astrónomos mundial, incluyendo a los astrónomos chilenos. Es por ello que un Observatorio Virtual Chileno (ChiVO, por sus siglas en inglés) es necesario y en la presente memoria se propone a ChiVO orientado a datos radioastronómicos de ALMA como primer acercamiento.

1.1 Definición del problema

Debido al rápido crecimiento del volumen de datos astronómicos, surgen 3 problemas de gran interés para la comunidad astronómica chilena. El primero es que se puedan buscar, encontrar y obtener datos desde múltiples fuentes a la vez, a pesar de que cada fuente puede requerir una manera de funcionar distinta dependiendo del instrumento que genere los datos. El segundo es que los datos deben ser accesibles en un tiempo razonable para los astrónomos. El tercero, es que se puedan hacer análisis sobre estos datos para que los astrónomos puedan hacer ciencia sobre éstos.

En el presente trabajo se busca diseñar una solución que permita resolver los tres problemas mencionados para datos de ALMA, identificando las necesidades de los radioastrónomos chilenos y proponiendo un modelo de datos que permita buscar y acceder a los datos que necesite el astrónomo. Así, facilitando el acceso a los datos se busca resolver en distintos grados los tres problemas.

1.1.1. Descripción del problema

El problema de los astrónomos chilenos a abarcar es encontrar datos y compararlos, ambos sobre muchos (y cada vez más) cantidades de datos provenientes de múltiples fuentes. Para ello hay que lidiar con los siguientes problemas:

Volumen de datos

Actualmente, el tamaño de los datos públicos astronómicos se estima en 1 PB, lo cual irá aumentando durante los años en el orden de 0.5 PB por año debido a que los

observatorios modernos almacenan los datos de las observaciones de manera digital, lo que facilita la obtención de información mediante análisis a partir de dichos datos.

En el futuro cercano habrá nuevos observatorios que generarán muchísimo más volumen de datos (ALMA, LSST, E-ELT) estimando que para el 2020 exista 60 PB de datos públicos, por lo que se hace necesario desde ya el diseño de sistemas que puedan soportar tales volúmenes y que cubran las necesidades de los astrónomos.

Los datos

Todo el volumen de datos mencionado anteriormente proviene de distintos observatorios y por lo tanto son representados, accedidos y analizados de distintas maneras ya que son requeridos por distintos campos dentro de la astronomía y por lo tanto, analizados de distintas maneras para comprobar o refutar distintas hipótesis. Todo lo anterior es un conjunto de problemas que sumados con la variedad de herramientas que se deben crear y mantener, generan una compleja problemática.

Chile

A pesar de que Chile es la capital mundial astronómica, Chile no cuenta con un Observatorio Virtual desaprovechando el gran potencial del país, ya que Chile cuenta con cerca del 10% de las observaciones realizadas en su territorio.

Observatorio Virtual

La solución a los problemas planteados anteriormente es la creación de un Observatorio Virtual Chileno, ya que logra la integración de datos, herramientas y servicios astronómicos a escala global de manera tal que facilita el acceso a la información por parte de los astrónomos y usuarios en general, lo cual es posible gracias a una alianza internacional de observatorios virtuales que es el núcleo de todo este ecosistema.

En el mundo ya se han creado OV que solucionan en parte los problemas presentados anteriormente sin embargo, en el caso de ALMA que genera datos de múltiples dimensiones (de ahora en adelante llamados cubos), no existe un OV que provea una

solución completa a los astrónomos. Además, los astrónomos chilenos obtienen los datos generados por ALMA desde los ARC, *ALMA Regional Centers*, que contienen una réplica de los datos de ALMA en el extranjero, provocando una demora en la entrega de los datos ya que estos salen de Chile para luego volver desde los ARC cada vez que un astrónomo chileno necesita datos.

Es por ello que este trabajo busca definir el diseño conceptual de un Observatorio Virtual especializado en datos generados por ALMA pensando en las necesidades de la comunidad astronómica chilena, o sea, un Chilean Virtual Observatory (ChiVO).

Debido a lo anterior, el objetivo general de esta memoria es

“Diseño conceptual de un Observatorio Virtual Astronómico para ALMA.”

Y los objetivos específicos, que una vez cumplidos logran el objetivo general son:

- Identificar los requerimientos para un OV chileno.
- Investigar y analizar estándares de IVOA.
- Investigar y analizar formatos para un OV.
- Investigar y analizar aspectos relevantes de ALMA.
- Definir un Modelo de Datos para un OV.

1.1.2. Alcance

Ya que el diseño de un OV requiere un gran conocimiento en informática tales como bases de datos, protocolos de transmisión de datos, entre otros y en radioastronomía orientada a ALMA, es necesario acotar el alcance de este trabajo de manera tal que sirva como base para la implementación factible del ChiVO.

Debido a la escasez de datos de ALMA en los OV, principalmente debido a su reciente inicio de operaciones y al interés en la comunidad chilena, se diseñará pensando en ALMA pero dejando espacio para todo tipo de datos al respetar los estándares y protocolos establecidos por IVOA.

Si bien ALMA aún no ha definido cómo ni en qué formato entregará sus datos, es seguro decir que harán al menos entrega de datos en formato FITS ya que es actualmente el formato más utilizado en la astronomía, además, es posible que también hagan entrega del ASDM, ambos formatos son entregados actualmente pero no se ha decidido por uno oficial. Es por ello que en esta memoria se tendrá en cuenta que cualquiera de estos dos formatos puede ser utilizado. Sin embargo, debido a la prematurez que se encuentran los proyectos que buscan analizar datos de ALMA, este trabajo se enfocará en la búsqueda de datos radioastronómicos de ALMA utilizando los estándares de IVOA actuales. Así, este trabajo se encuentra suficientemente acotado para el tiempo de una memoria de pregrado y la posibilidad de una implementación a corto plazo.

1.1.3. Metodología

Para poder resolver el problema identificado para los usuarios, los radioastrónomos chilenos, se debe seguir una metodología que busque una solución siempre en contacto con los usuarios finales y que a medida que se vayan concretando pasos, éstos se validen. La siguiente lista indica los pasos que describen la metodología a utilizar:

1. Identificación de *stakeholders*: La identificación de *stakeholders* se realizará sumergiéndose en el área de la radioastronomía, Observatorios Virtuales y especialmente, ALMA. Para ello se estará activamente presente en el proyecto de Fondef que construye las bases de ChiVO, se asistirá a conferencias y *workshops* de astroinformática y se tendrán entrevistas regulares con expertos en las tres áreas mencionadas.

2. Identificación de problemas a resolver: Para identificar los problemas, se realizarán las entrevistas antes mencionadas a todos los *stakeholders*, incluyendo radioastrónomos, científicos con experiencia en Observatorios Virtuales extranjeros y

científicos de ALMA. Además, en esta fase y en la anterior, el estudio del estado del arte será fundamental.

3. Análisis de requerimientos: Luego de los pasos anteriores, se obtendrá una visión clara del panorama chileno e internacional. A continuación se realizará una lista de los requerimientos de búsquedas en los Observatorios Virtuales en conjunto y con la posterior validación de radioastrónomos chilenos.

4. Definición de requerimientos y casos de uso: A partir de los requerimientos se definirán los actores, se identificarán los casos de uso correspondientes para cada requerimientos en detalle. Finalmente se validarán y verificarán con los *stakeholders*.

5. Modelo de Datos: Como parte del diseño y a partir de lo anterior, se reconocerán los estándares de IVOA necesarios para poder concretar los requerimientos de ChiVO, en base a eso se diseñará el modelo de datos.

En el capítulo 2 se describe el estudio de estado del arte sobre Observatorios Virtuales, astronomía en Chile y otros temas relevantes, en el capítulo 3 se encuentra el resultado de la metodología especificada incluyendo requerimientos, casos de uso y modelos de datos de ChiVO, y en el capítulo 4 se termina con las conclusiones.

Capítulo 2: Estado del Arte

El uso de OV ha ido incrementándose durante los últimos años así como la creación de estos sistemas, sobre todo desde la formación de la alianza internacional de observatorios virtuales (IVOA) en el 2002.

A continuación, se menciona el trabajo realizado por esta alianza internacional, el trabajo realizado por algunos de los OV integrantes, qué es lo que Chile ofrece actualmente junto con lo necesario para la creación del ChiVO, los últimos avances en la radioastronomía introducidos por ALMA. También se mencionan algunos servicios astronómicos que son de gran utilidad para el proyecto y los formatos de archivos más utilizados para almacenar datos astronómicos en la actualidad.

2.1. La Alianza Internacional de Observatorios Virtuales (IVOA)

Uno de los principales objetivos de un Observatorio Virtual es facilitar la búsqueda de los datos que requiera el astrónomo, permitiéndole a éste una búsqueda en todos los datos de todos los instrumentos existentes. Para ello, sería muy costoso que un organismo cree una forma de acceso única para cada instrumento, es por ello que es necesario estandarizar la forma en que los datos son publicados.

Es por ello que en el 2002 se creó una alianza internacional de observatorios virtuales (IVOA) con la misión de “facilitar la coordinación y colaboración internacional necesaria para el desarrollo e implementación de herramientas, sistemas y estructuras organizacionales necesarias para habilitar la utilización internacional de archivos astronómicos como un observatorio virtual integral e interoperativo” [4]. El IVOA ha crecido bastante desde sus inicios ya que ahora cuenta con más de 20 miembros y aún más próximos a unirse, incluyendo ChiVO.

El IVOA, con el objetivo de lograr su misión, desarrolla estándares junto a la comunidad y promueve el uso de éstos entre sus integrantes. Para ello, se definió una

arquitectura identificando los componentes principales de un OV. Para cada uno de estos componentes se desarrollan estándares que son definidos por grupos de trabajos que siguen un proceso modelado por el *World Wide Web Consortium*, donde cada grupo de trabajo genera borradores de futuros estándares que a medida que son aprobados por la comunidad, pasan a ser recomendaciones propuestas y finalmente, recomendaciones. Quiénes respeten las recomendaciones son compatibles con todo el ecosistema de los OV, en todo este proceso se promueve que expertos en cada área de la comunidad participen. IVOA también tiene grupos dedicados a recibir retroalimentación de las experiencias que han tenido cada uno de sus miembros con el uso de los estándares, lo que permite que estos estándares se acomoden lo máximo posible a las necesidades reales de la comunidad astronómica.

Cómo se menciona en el sitio web del IVOA [8], un OV tiene dos componentes principales:

- Encontrar qué datos están disponibles.
- Obtener acceso a dichos datos a través de simples consultas.

La idea es construir estos componentes como servicios web que ofrezcan funcionalidad remota de tal manera que esté disponible al usuario como si fuese local en su computador. Estos servicios ya llevan una descripción de sí mismos lo que va acorde a un servicio web además, los proyectos de OV han estado y seguirán agregando más información específica de astronomía a esta descripción, lo que facilita la transmisión de los datos. Estas descripciones de servicios y sus conjuntos de datos, están disponibles a través de registros, los cuales cumplen ciertos estándares de manera tal que sean compatibles con los OV. Estas implementaciones de registros son construidas independientes entre sí, pero se comunican entre ellas para entregar una vista unificada, lo cual se asemeja a un sistema existente en el Internet, los DNS (*Domain Name System*) cuya función principal es traducir nombres de dominio de un sitio web a una IP.

Como se mencionó anteriormente, la arquitectura de un OV es orientada al servicio, lo que significa que los componentes del sistema son definidos por la naturaleza de las peticiones y respuestas a estos servicios. Debido a esto, la descripción de cada servicio está basada en la elección de los protocolos para las peticiones y respuestas, en vez de las clases o métodos. Cada servicio es autónomo, y sus límites están bien definidos. Los servicios son intrínsecamente distribuidos, de manera tal que puedan ser implementados en cualquier máquina que parezca óptima, manteniendo la vista unificada.

Los datos requeridos por el usuario son enviados entre los servicios en dos formatos básicos: FITS, un estándar astronómico desde hace muchos años; y XML, un sintaxis estándar para codificar información. En este último caso, el proceso de IVOA permite gran flexibilidad acorde a las necesidades que se tengan, ya que permite una propuesta de esquema convertirse en un nuevo estándar a través de un proceso de comunidad bien definido; ejemplos exitosos son: un nuevo formato para representar datos tabulares ricos en metadatos que separa los datos de los metadatos haciendo más fácil las búsquedas su nombre es VOTable, otro ejemplo es un esquema XML que contiene metadatos que describen entidades que tienen registros compatibles con OV llamado VOResource [8].

Así los servicios compatibles con IVOA se construyen para intercambiar mensajes que pueden ser documentos y diccionarios en formato XML (conjuntos de pares palabra-clave), así como formatos tradicionales binarios tales como archivos FITS. Un servicio compatible con IVOA se define por la naturaleza de estos mensajes. La comunidad de proveedores de datos es promovida a implementar dichos servicios, y la comunidad de consumidores de datos es promovida a construir portales que utilizan estos servicios, generando así un ecosistema construido por y para la comunidad.

En la figura 2.1 se muestran los componentes esenciales de la arquitectura de un OV:

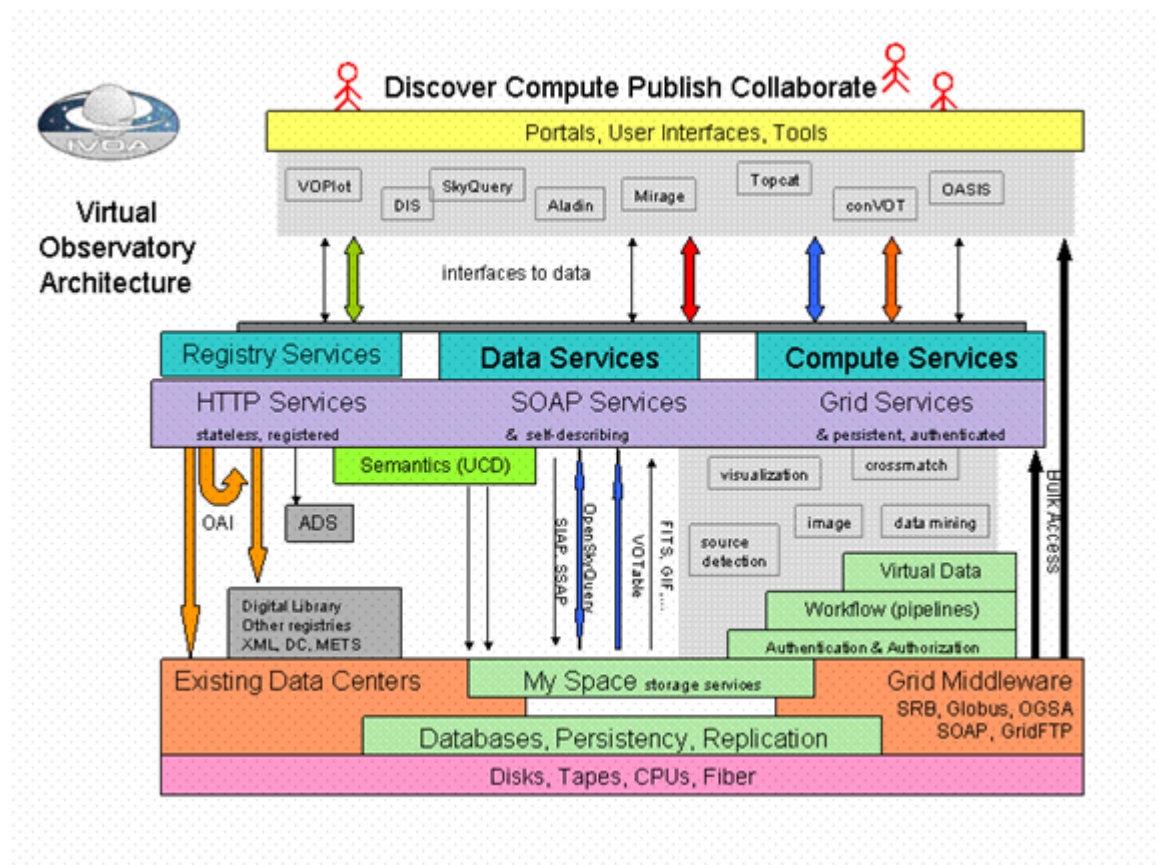


Figura 2.1. Arquitectura de un OV según IVOA

A continuación, se describen los componentes principales de la Figura 2.1.

2.1.1 Sección superior arquitectura IVOA

El objetivo principal de un Observatorio Virtual se muestra en la parte superior de la Figura 2.1: mejorar y unificar el acceso a los datos astronómicos y servicios tanto para los astrónomos profesionales, como también para el público en general. La barra superior de la figura representa este objetivo: el descubrimiento de datos y servicios, la reformulación y el análisis de esos datos a través de la computación, la edición y difusión de los resultados, y el aumento de la producción científica a través de la

colaboración. Debajo de la barra amarilla, se encuentra un rectángulo gris que contiene ejemplos de portales y herramientas que cumplen el objetivo de la barra superior.

Las flechas verticales de distintos colores representan los diferentes tipos de servicios y formatos XML por el que estos portales y herramientas interactúan con los servicios compatibles con IVOA. En la arquitectura IVOA, se han dividido los servicios disponibles en tres grandes categorías:

- Servicios de datos (*Data Services*), para servicios relativamente simples que proporcionan acceso a los datos,
- Servicios de cómputo (*Compute Services*), donde el énfasis está en la computación y la agrupación de datos.
- Servicios de registro (*Registry Services*), para permitir a servicios y otras entidades ser publicadas y descubiertas,

Estos servicios se llevan a cabo en varios niveles de sofisticación, desde un texto con una petición-respuesta, a un servicio autenticado y auto descriptivo que utiliza computación de alto rendimiento para construir una respuesta estructurada a partir de una solicitud estructurada. En los OV, se pretende que los servicios puedan ser utilizados no sólo individualmente, sino también concatenados en un *workflow*, donde la salida de uno es la entrada de otro [8].

2.1.2. Sección intermedia arquitectura IVOA

Los servicios de registro facilitan la publicación y descubrimiento de servicios. Si un proveedor de datos pone en línea un nuevo conjunto de datos, con un servicio para acceder a éste, el siguiente paso sería la publicación de este hecho a un registro compatible con los OV. Una forma de hacer esto es rellenar formularios donde se indica quién, dónde, cómo y para qué es el servicio. A su debido tiempo, los registros

se cosecharán entre sí (copiando nuevos registros) de tal manera que el nuevo conjunto de datos con su respectivo servicio serán conocidos para otros registros del OV. Cuando otra persona busque un registro (palabra clave, autor, área, longitud de onda, etc.), descubrirá los servicios publicados. De esta manera, los OV mejoran la difusión de información a un sistema más eficiente e igualitario.

En la arquitectura de un OV, no hay nadie que decida qué datos son buenos y cuales malos, (aunque los registros individuales pueden imponer criterios si así lo desean). En su lugar, IVOA espera de los buenos datos surgir de manera natural, como sucede en el *World Wide Web*. Hay que notar que si bien la web no tiene restricciones de publicación, sigue siendo un recurso de enorme utilidad, y se espera que el mismo paradigma hará los registros de OV útiles [8].

Continuando con los registros, cada uno tiene tres tipos de interfaz: Publicación, consulta y cosecha. La gente puede publicar a un registro completando formularios web en un portal web, definiendo así los servicios, las colecciones de datos, proyectos, organizaciones y otras entidades. El registro también puede aceptar consultas en uno o más lenguajes (por ejemplo, un lenguaje de consulta estándar del IVOA), de tal manera de descubrir entidades que cumplen los criterios especificados. La tercera interfaz, la cosecha, permite a los registros intercambiar información entre sí, de modo que una consulta que se ejecuta en un registro puede descubrir un recurso que fue publicado en otro.

Los servicios de registros esperan poder etiquetar cada recurso del OV a través de un identificador universal, que puede ser reconocido por la cadena inicial ivo://. Los recursos pueden contener enlaces a recursos relacionados, así como enlaces externos a la literatura, especialmente al *Astronomical Data System*. La arquitectura de registro del IVOA cumple con los estándares de bibliotecas digitales para la recolección de metadatos y esquemas de metadatos, con el objetivo de que los

recursos que sean compatibles con los estándares del IVOA puedan aparecer como parte de cualquier biblioteca universitaria.

Los servicios de datos van desde simples a sofisticado, y retornan datos tabulares, imágenes u otros. Al nivel más simple la solicitud es un cono en el cielo (coordenadas y radio angular, *conesearch*), y la respuesta es una lista de objetos, cada uno de ellos con una posición dentro del cono sobre el cielo. Servicios similares como el servicio para acceso de imágenes a través de SIAP (*Simple Image Access Protocol*) o el servicio que da acceso a espectros asociados a regiones del cielo SSAP (*Simple Spectral Access Protocol*), son servicios que también pueden ser capaces de realizar consultas sobre otros parámetros de los objetos.

También hay servicios de datos que permiten la consulta de una base de datos relacional o una aglomeración de bases de datos como el *Open SkyQuery*.

Los miembros de la IVOA están colaborando con grupos de TI que están desarrollando el *workflow software*, es decir, un conjunto enlazado de servicios distribuidos con un paradigma *workflow*. El objetivo es reutilizar componentes de los servicios para construir aplicaciones complejas, donde los servicios están aislados unos de otros a través de protocolos bien definidos, y por lo tanto más fáciles de mantener y depurar. Los miembros del IVOA también esperan utilizar estos *workflows* en el contexto de datos virtuales, en el sentido de que un dato como producto se genera dinámicamente sólo cuando es necesario, a pesar de tener un caché de datos precalculados que puede ser utilizado cuando se requiera.

2.1.3. Sección inferior arquitectura IVOA

En el diagrama anterior, la capa más baja es el hardware actual, pero sobre éstos están los centros de datos existentes, que implementan y/o depuran los servicios estándares del IVOA. El *middleware Grid* es utilizado para computación de alto

rendimiento, transferencia de datos, autenticación y ambientes de servicio. Otros componentes de software incluyen bases de datos relacionales, servicios para replicar las colecciones de uso frecuente, y las *Grid* de datos para la gestión de colecciones distribuidas.

Una parte vital de la arquitectura del IVOA es que los usuarios puedan almacenar datos dentro del OV, esto es llamado *MySpace*. *MySpace* almacena archivos y tablas bases de datos obtenidas desde algún OV u otra fuente y los hace accesibles para todos los servicios dentro del ecosistema OV, evitando la necesidad de recuperar los resultados del computador del usuario reduciendo así tiempos de transferencia de datos. El uso de *MySpace* establece derechos de acceso y privacidad sobre resultados intermedios y permite a los usuarios gestionar su almacenamiento remotamente [8].

2.2 Estándares de IVOA

2.2.1 *Observation Core*

El estándar de modelo de datos *Observation Core* busca representar la mayor cantidad de observaciones realizadas por instrumentos astronómicos en todo el mundo, permitiendo así la búsqueda y el descubrimiento sobre éstos. Es un estándar de IVOA que se encuentra en estado de recomendación

La posibilidad de plantear una consulta científica a múltiples fuentes al mismo tiempo es un caso de uso fundamental para el Observatorio Virtual. Al proveer un simple protocolo estándar se incrementan las posibilidades de que la mayoría de los proveedores de datos en astronomía sean capaces de implementar el protocolo, permitiendo de este modo el descubrimiento de datos para casi todas las observaciones astronómicas registradas.

Es por ello que *Observation Core* es el núcleo de cualquier observación.

2.2.1.1. Principales conceptos del modelo de datos *Observation Core*

En general, una observación puede estar compuesta por múltiples productos de datos individuales. En estos casos todos los productos de datos que comprenden una observación deberían compartir el mismo identificador de observación (*obs_id*). La forma del *string obs_id* queda en los proveedores de datos con tal que identifiquen únicamente una observación dentro del archivo.

Las columnas representadas en el modelo de datos de *Observation Core* están divididas en dos conjuntos. Aquellos necesarios para apoyar los casos de uso provistos por IVOA y otros que son generalmente útiles para describir datos, pero no inmediatamente requeridos para apoyar los casos de uso. A continuación, en la Tabla 2.1 se describen sólo los del primer grupo ya que es lo mínimo necesario para ser totalmente compatible con los estándares de IVOA.

Tabla 2.1. Campos obligatorios del estándar de IVOA *Observation Core*.

Nombre de la Columna	Unidad	Tipo	Descripción
dataprodut_type	Sin unidad	string	Tipo de producto de los datos lógicos (imagen, cube, etc.)
calib_level	Sin unidad	enum integer	Nivel de calibración {0, 1, 2, 3}
obs_collection	Sin unidad	string	Nombre de la colección de datos
obs_id	Sin unidad	string	ID de la observación
obs_publisher_did	Sin unidad	string	Identificador del conjunto de datos dado por el que publica
access_url	Sin unidad	string	URL para acceder (descargar) el conjunto de datos
access_format	Sin unidad	string	Formato del contenido del archivo
access_estsize	kbyte	integer	Tamaño estimado del conjunto de datos en kilo bytes
target_name	Sin unidad	string	Objeto astronómico observado, en caso de que haya
s_ra	grado	double	Ascensión central derecha, ICRS
s_dec	grado	double	Declinación central, ICRS
s_fov	grado	double	Diámetro (límites) de la región cubierta
s_region	Sin unidad	CoordArea	Región cubierta como es especificado en STC o ADQL
s_resolution	arcseg	float	Resolución espacial de los datos como FWHM
t_min	fecha	double	Tiempo de inicio en MJD
t_max	fecha	double	Tiempo de fin en MJD
t_exptime	segundo	float	Tiempo total de exposición
t_resolution	segundo	float	FWHM de la resolución temporal
em_min	Hz	double	Inicio en coordenadas espectrales
em_max	Hz	double	Fin en coordenadas espectrales
em_res_power	Sin unidad	double	Poder de resolución espectral
o_ucd	Sin unidad	string	UCD de lo observable (ej. phot.flux.density)
pol_states	Sin unidad	string	Lista de los estados de la polarización (NULL si no aplica)
facility_name	Sin unidad	string	Nombre de la instalación utilizada para esta observación
instrument_name	Sin unidad	string	Nombre del instrumento utilizado para esta observación

Una descripción más detallada de cada columna puede ser encontrada en el Anexo B.

2.2.1.2. Implementación del modelo de datos *Observation Core* en un servicio de acceso

Para poder acceder a los datos contenidos en el modelo de datos *Observation Core*, éste debe ser implementado en conjunto con un servicio de acceso a tablas, esto está descrito por el estándar *Table Access Protocol* (TAP) de manera que todas las consultas válidas puedan ser ejecutadas sin cambios en cualquier servicio que haya implementado el modelo. Se permiten columnas adicionales siempre y cuando se especifiquen todas las columnas obligatorias.

El modelo se especifica explícitamente mediante tablas de una base de datos relacional, en el contexto del TAP esto significa que se especifican tablas lógicas como las expuestas en *TAP_SCHEMA* (donde las tablas y columnas expuestas por el servicio son descritas). Esto no implica que la base de datos tenga la misma estructura pero en la mayoría de los casos la relación entre la descripción *TAP_SCHEMA* y las tablas de la base de datos es sencilla.

2.2.2. Modelo de datos de caracterización (CharDM)

El estándar de modelo de datos *Characterisation* es aplicable a datos observados o simulados, pero no está diseñado para catálogos, tales como listas de propiedades derivadas o fuentes. El modelo está destinado a describir:

- Una observación simple;
- Una colección de datos;
- El parámetro espacio usado por una herramienta o paquete accedido a través del OV.

El estándar de modelo de datos *Characterisation* satisface dos requerimientos principales el de descubrimiento de datos y el de análisis/procesamiento de datos.

2.2.2.1. Principales conceptos del modelo de datos *Characterisation*

Considerando una típica observación astronómica, se identifican con cuatro Propiedades (*Properties*) principales y algunas subpropiedades:

- Cobertura (*Coverage*): describe la dirección a la que estaba apuntando el telescopio, en cuales longitudes de onda y cuando; y/o la región cubierta por cada eje. Está descrito en niveles crecientes de detalle por:
 - Ubicación (*Location*).
 - Límites (*Bounds*).
 - Soporte (*Support*).
 - Sensibilidad (*Sensitivity*).
 - Factor de relleno (*Filling factor*) si el soporte no está definido precisamente.
- Precisión de Muestreo (*Sampling Precision*): describe los intervalos de muestreo de cada eje.
- Resolución (*Resolution*): describe la resolución física efectiva.
- Precisión (*Accuracy*): describe la precisión de la medición.

2.2.2.2. Ejes y sus atributos

Las dimensiones físicas para los datos están descritas por ejes tales como: espacial, espectral, tiempo, velocidad, visibilidad, polarización, observable. Se recomienda que los proveedores de datos usen estos nombres, pero no es obligatorio (nombres FITS pueden ser usados). Cada eje contiene los conceptos mencionados anteriormente, pero no necesariamente todos. El proveedor de datos deberá proporcionar un UCD y las unidades para cada eje. No hay límite en el número de ejes presentados y pueden ser dependientes o superpuestos (por ejemplo un eje de frecuencia y dos ejes de velocidad representando dos moléculas separadas con transiciones a frecuencias similares).

Algunos ejes no pueden ser representados por los datos, pero son implícitos, presentados solo con una palabra clave en la cabecera o en otro lugar. Por ejemplo, una simple imagen del cielo 2D tiene ejes coordenados celestiales, pero los ejes espectral y tiempo no pueden ser presentados en el arreglo de datos principal a través

de la observación usando una integración finita de tiempo y bandas de longitud de onda.

Cada eje contiene banderas o *flags* que son usadas para indicar propiedades de calificación booleanas y otras. Las *flags* presentes en un ejes son si el eje es independiente, el estado de calibración y si los datos están bajo muestreo.

Además, cada eje contiene la precisión que caracteriza alguna incerteza asociada con cada uno.

2.2.2.3. Cobertura (*Coverage*)

Tiene distintos niveles de profundidad, proporcionando un rango de detalle dependiendo de las necesidades de algún usuario/desarrollador. La cobertura es descrita en cuatro capas las cuales entregan una visión jerárquica de detalle incremental.

1. Ubicación (*Location*): el elemento más simple es la posición de un punto en un parámetro espacio N-dimensional. Una definición precisa no es requerida, pero *Location* puede servir como un valor de referencia u origen de coordenadas en cuadros sin una posición absoluta.
2. Límites (*Bounds*): El siguiente nivel de descripción es el *SensitivityBounds*. Un rango en cada parámetro proporcionando el límite inferior y límite superior de una “caja” N-dimensional. Los límites garantizan encerrar todos los datos válidos, aunque podría haber regiones del borde en las que no los hay.
3. Soporte (*Support*): describe cuantitativamente el subconjunto de espacio, tiempo, frecuencia y otros dominios en los cuales hay datos válidos (de acuerdo a un criterio de calidad especificado). Puede incluir uno o varios rangos en cada eje.
4. Sensibilidad (*Sensitivity*): provee valores numéricos indicando la variación de la función de respuesta en cada eje.

Los límites pueden contener también un sub nivel de factor de llenado (*Filling Factor sub-level*), el cual entrega la fracción útil de los límites en algún eje. No es recomendable usar si hay muchas interrupciones. Llenados muy regulares pueden ser descritos por *Sampling*.

2.2.2.4. Resolución (*Resolution*) y Precisión de Muestreo

- Resolución (*Resolution*): es usualmente el mínimo intervalo independiente de medición en cada eje.
- Muestreo (*Sampling*): también llamado pixelación o precisión, describe el truncamiento de valores de datos como parte de la adquisición o procesamiento de datos. Si el muestreo no es lineal, puede ser necesaria una simplificación entregando valores límite de una “*characteristic sampling precision*”.

2.2.2.5. El Modelo

Se usan diagramas UML para describir la organización de los metadatos de Characterisation siguiendo la perspectiva Propiedades/Ejes/Niveles (*Properties/Axis/Levels perspective*). El modelo ofrece diferentes vistas de los conceptos de caracterización. Cada propiedad está asociada a un tipo de eje (*AxisType*), que representa el eje por medio del cual la propiedad es evaluada.

2.2.2.6. Ejes en detalle

Toda la información relacionada con un eje es reunida dentro de la clase *CharacterisationAxis*. Contiene el nombre del eje, unidades, UCD, sistema de coordenadas STC, ubicación del observatorio. Además, cada eje puede incluir el número de contenedores (*bins*) presentes en el eje y *flags* que indiquen el estado de la calibración, propiedades de independencia y muestreo del eje.

Si un objeto de un nivel inferior necesita tener su propia descripción de ejes puede ser definida localmente sobre escribiendo el objeto de alto nivel *CharacterisationAxis*. La redefinición puede ser parcial.

Estado de calibración

Characterisation provee un *flag* de estado por cada eje, así el usuario puede insistir en datos calibrados solo cuando es necesario. *CharacterisationStatus* es entregado por separado para cada tipo de eje de caracterización y puede ser:

- *UNCALIBRATED*: en unidades que no pueden ser directamente comparados con otros datos.
- *CALIBRATED*: en unidades confiables u otras aceptadas como magnitudes.
- *RELATIVE*: calibrado dentro de un factor constante, el cual no es precisamente conocido.
- *NORMALIZED*: datos adimensionales, divididos por otro conjunto de datos (o un extremo local).

Estado de muestreo

- *Undersampling*: TRUE si el periodo de precisión del muestreo es grande comparado con la resolución y la precisión de un valor de datos es limitado para el muestreo; FALSE si el periodo de precisión del muestreo es pequeño comparado con la resolución y la precisión es limitada por la resolución.
- Muestreo regular (*Regular sampling*): TRUE si la pixelación o agrupación (*binning*) es casi lineal con respecto al eje de las coordenadas globales; FALSE si introduciría un error significativo con respecto a otras incertidumbres.
- El número total de muestras a través de cada eje puede ser entregada, normalmente usado por múltiples muestreos regulares.

2.2.2.7. Errores en Characterisation (*Accuracy*)

Los errores pueden estar en las unidades de los ejes o estar representados por indicadores de calidad (*quality flags*). Estos errores son producidos en un objeto *Accuracy*. *Accuracy* soporta múltiples niveles de descripción, análogo a *Coverage*. La incertidumbre en la posición o medición de algún eje puede ser descrita por un valor típico, por los límites en un rango de errores y/o por valores de error muy detallados para cada elemento de muestreo.

2.2.2.8. Implementando el modelo usando elementos STC

El estándar *Space-Time Coordinate* abarca la descripción de la mayor parte de los ejemplos de ejes de *Characterisation*, excepto el *Observable*. *Sensitivity* es la única propiedad ausente en STC. No es posible utilizar la estructura completa de STC porque no tiene la flexibilidad para entregar una alternativa a los esquemas multicapas. Se utilizan objetos de STC de nivel intermedio como bloques de construcción del modelo *Characterisation*.

- El objeto *STC:AstroCoordSystem* es necesario como referencia para el eje *Coverage*. Subestructuras de STC pueden ser usadas como:
- *Location* implementa *STC:AstroCoords*.
- *Bounds* implementa tipos básicos de STC, algunos elementos de *STC:Interval* y *STC:Coords* en una estructura similar a *STC:AstroCoordArea*.
- *Support* usa *STC:AstroCoordArea*.
- *Resolution* *RefVal* puede ser implementada vía tipos idóneos usando elementos *STC:CResolution*.
- *SamplingPeriod* y *SampleExent* encapsulan elementos de *STC:CPixSize*.

2.2.3. TAP

El servicio *Table Access Protocol*, establecido por el IVOA, especifica el intercambio de mensajes entre el cliente y un servicio para el acceso a datos astronómicos, a través de métodos para realizar consultas y recibir los resultados obtenidos a través de éstas.

El uso de TAP se basa en una estructura de árbol sobre una URL. Para acceder a una característica de un servicio TAP de observatorio virtual formando una URL con la URL base del sitio, luego “TAP” y finalmente la característica que se desee acceder¹.

¹ [www.example.com/TAP/\[CARACTERÍSTICA\]](http://www.example.com/TAP/[CARACTERÍSTICA])

Las consultas se implementan bajo los protocolos POST (envío de datos para ser procesados en el recurso especificado) y GET (Petición de datos de un recurso especificado). Existen dos tipos de consultas. El primer tipo de consulta es la sincrónica, se debe utilizar preferentemente para obtener pequeñas cantidades información, ya que solo responde una tabla con resultados y que se ejecutan inmediatamente una vez que la petición es recibida, y tiene un tiempo definido para responder, entregando los resultados en HTTP *response* o HTTP *redirect*. Si los resultados no están listos después del tiempo correspondiente, la consulta falla y no se entregan datos. Su implementación es obligatoria. La URL a armar es el URL base del sitio, luego “TAP” y la característica que en este caso es “sync”².

Para consultas asincrónicas se deben armar a través de HTTP *exchanges*, basados en UWS (*Universal Worker Service*), para manipular y manejar los trabajos que serán ejecutados por las consultas asincrónicas, indican el estado del servicio y de la ejecución del trabajo, este tipo de consultas se ejecutan cuando se tienen todos los parámetros. La URL se arma partiendo con la URL base del sitio, luego añadiendo “async”, luego el trabajo a realizar y finalmente la capacidad a utilizar³. Entre las capacidades disponibles están *availability* que indica si el servicio esta disponible, *capability* que muestra las capacidades del servicio y *tables* que indica los metadatos de las tablas utilizadas.

Tanto para las peticiones sincrónicas y asincrónicas se especifican los parámetros que se deben aceptar, estos parámetros son enviados a través de HTTP POST, notar que no todas las combinaciones de parámetros son posibles, para consultas falsas o erróneas se debe ignorar el parámetro falso y responder de forma normal. *REQUEST*, distingue la operación, si no es incluido se debe enviar un error, los valores posibles son *doQuery* y *getCapabilities*. *Version* indica la versión de TAP que se quiere utilizar, *LANG* especifica el lenguaje utilizado en la consulta (ADQL, PQL u otro),

² www.example.com/TAP/sync/

³ [www.example.com/TAP/async/\[JOB\]/\[CAPABILITY\]](http://www.example.com/TAP/async/[JOB]/[CAPABILITY])

QUERY se especifica la consulta según el lenguaje de consultas especificado, FORMAT especifica el formato (MIME) que estará representado el resultado (VOTable, CSV, TSV, FITS, *text*, HTML), MAXREC indica el número de resultados que se quieren ser devueltos, si no es indicado debe existir un número predeterminado y si es 0, se deben devolver los metadatos de las tablas, RUNID identifica el *Job* o servicio, UPLOAD utilizado para subir tablas.

Se podrán subir tablas en el formato VOTable, se debe especificar el nombre según el estándar de ADQL, para acceder a las tablas se refiere a TAP UPLOAD.[nombre de la tabla], estas tablas no se deben mostrar en el TAP SCHEMA Todas las consultas sincrónicas y asincrónicas deben especificar REQUEST=doQuery y LANG.

Las tablas dentro de la base de datos están bajo siguiente esquema de nombres:

[[catalog name.]schema.]table_name

, donde *catalog name* corresponde al nombre del catálogo o base de datos, *schema* se especifica el esquema de datos y *table name* el nombre de la tabla. Se debe tener un *TAP SCHEMA* donde especifiquen los metadatos de las tablas de TAP. Las columnas que deben estar presentes para ser consultadas son SCHEMA.schemas, TAP SCHEMA.tables, TAP SCHEMA.Columns, TAP SCHEMA.keys y TAP SCHEMA.key, se podrá acceder a esta información a través de una consulta con MAXREC=0.

Para cualquier tipo de consultas existen lenguajes de consulta determinadas, ADQL es obligatorio implementarlo, se debe usar comúnmente para acceder a una o más tablas, incluyendo los metadatos, PQL es opcional y es utilizado para acceder a una sola tabla del servicio, o sus metadatos, se puede dar soporte a otro tipo de lenguajes de consultas, los valores de los parámetros en ADQL están basados en STC-S, para distintas funciones como POINT, CIRCLE, BOX, POLYGON.

TAP es un servicio web que esta basado en las reglas de uso de HTTP IETF RFC 2616 y debe soportar GET y POST, la especificación de la URL IETF RFC 2396. Los parámetros de /async y /sync son enviados por POST.

Las respuestas deben ser del tipo especificado en FORMAT, si no se especifica, predeterminadamente debe ser VOTable, y la representación dependerá del lenguaje de la consulta, el orden de las columnas en los resultados deben corresponder al orden en el parámetro SELECT.

Cuando se detecte un error se debe enviar documento de error (VOTable) y código HTTP-status correspondiente, los errores pueden ser generados cuando, por ejemplo, se intenta acceder en /async a un trabajo (*job*, *UWS joblist*) que no existe, en cuyo caso se debe enviar un HTTP 404, y cuando se intente acceder a un recurso al que no se tiene autorización HTTP 401.

La publicación de un OV debe ser registrada en IVOA, describiendo la identidad y las capacidades del servicio, se debe tener un documento estructurado como VOResource, del subtipo catálogo de servicio (*CatalogService*), indicando la interfaz, las funciones del TAP y capacidades extras del VO, así como la URL de TAP.

2.2.3.1. Implementaciones

Existen varias implementaciones de TAP, a través de un servicio Web, hasta programas de líneas de comando, TAPsh es un TAP de línea de comandos para acceder a los datos y servicios a través de consola, esta hecho en java y python. Muy similar es TOPCAT, que permite obtener tablas VOTable o FITS y analizarlas a través de una interfaz gráfica. GAVO, el observatorio virtual alemán, implementó una servicio TAP web, se pueden hacer consultas asincrónicas y sincrónicas.

2.3. Observatorios Virtuales y el mundo

Todos los miembros de IVOA tienen proyectos que son o están siendo realizados con tal de contribuir al Observatorio Virtual. Los miembros actuales de IVOA son los que aparecen en la Tabla 2.2.

Tabla 2.2. Miembros de IVOA.

Proyecto	Sitio Web
Argentina Virtual Observatory	http://nova.conicet.gov.ar/
Armenian Virtual Observatory	http://www.aras.am/Arvo/arvo.htm
AstroGrid	http://www.astrogrid.org/
Australian Virtual Observatory	http://aus-vo.org.au/
Brazilian Virtual Observatory	http://www.lna.br/bravo/
Canadian Virtual Observatory	http://www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cvo/
Chilean Virtual Observatory	http://www.chivo.cl/
Chinese Virtual Observatory	http://www.china-vo.org/
European Space Agency	http://www.sciops.esa.int/index.php?project=ESAVO
European Virtual Observatory	http://www.euro-vo.org/
German Astrophysical Virtual Observatory	http://www.g-vo.org/
Hungarian Virtual Observatory	http://hvo.elte.hu/en/
Italian Virtual Observatory	http://vobs.astro.it/
Japanese Virtual Observatory	http://jvo.nao.ac.jp/
Observatoire Virtual France	http://www.france-vo.org/
Russian Virtual Observatory	http://www.inasan.rssi.ru/eng/rvo/
South Africa Virtual Observatory	http://www.sao.ac.za/resources/virtual-observatory/
Spanish Virtual Observatory	http://svo.cab.inta-csic.es/
Ukrainian Virtual Observatory	http://www.ukr-vo.org/
US Virtual Astronomical Observatory	http://www.usvao.org/
Virtual Observatory India	http://vo.iucaa.ernet.in/~voi/

En la Tabla 2.2 se pueden apreciar los 21 miembros de IVOA, Chile se incorpora este año 2013. Los miembros de IVOA son agrupaciones nacionales de desarrolladores y proveedores con un fuerte compromiso a contribuir al ecosistema OV, a través de cambios fundamentales en la manera que fueron diseñados y operan sus Observatorios nacionales. La ventaja de ser miembro es el acceso a la experiencia de otros miembros junto a científicos de renombre mundial [9].

En la Figura 2.2 se indica el grado de alcance que podrían tener los OV mostrando el porcentaje de observatorios mundiales que están en países que son miembros de IVOA.



Figura 2.2. Gráfico de observatorios miembros de IVOA.

Cabe destacar que los 20 países con mayor cantidad de observatorios del mundo, tienen o son partes de algún Observatorio Virtual, lo que se traduce en que el 73.9% de los observatorios científicos alrededor del mundo se encuentran en países miembros de IVOA [10], como se puede apreciar en la Figura 2.2. En la Figura 2.3 se muestra la cobertura del espectro electromagnético por los observatorios más importantes y por los observatorios virtuales.

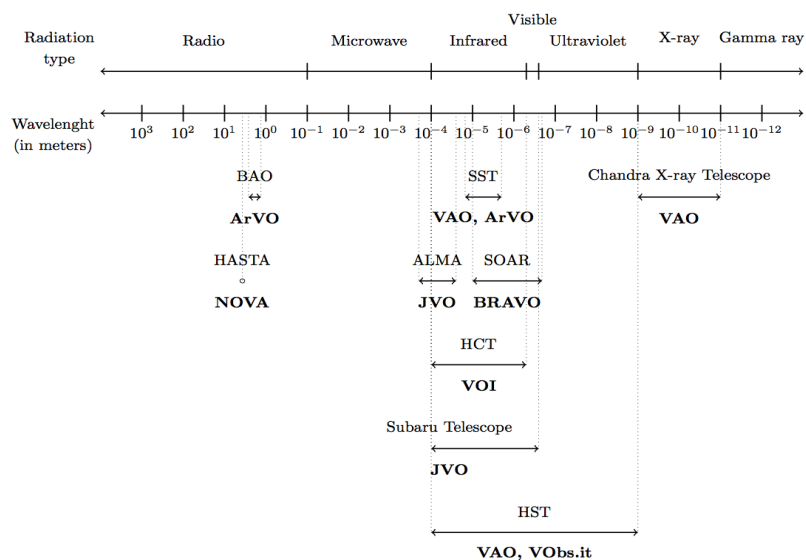


Figura 2.3. Cobertura del espectro electromagnético por distintos instrumentos y OV.

En la Figura 2.3 se muestra la cobertura que tienen algunos de los instrumentos más importantes en el espectro electromagnético, en el caso de ALMA el único OV que lo cubre es el Observatorio Virtual Japonés (JVO).

De los miembros de IVOA no todos proveen servicios de OV ya que no todos han implementado los estándares de IVOA o no ha sido publicada su implementación. Sin embargo, la mayoría de los miembros de IVOA tienen proyectos realizados o en curso lo cual es un gran logro ya que evidencia la seriedad de IVOA y del impacto que ya está generando en la astronomía y en la humanidad. En el Anexo C se encuentra una lista completa con todos los OV y proyectos activos/en progreso/anunciados.

2.4. Estado del Arte Nacional

Actualmente, Chile tiene el potencial de ser la capital astronómica del mundo, oportunidad que ha desaprovechado durante años. Es por ello que estos últimos años se han hecho una gran cantidad de esfuerzos para promover el estudio de la astronomía ya que para el 2020, con el E-ELT (*European Extremely Large Telescope*) se espera que Chile albergue sobre el 70% del total de la infraestructura astronómica mundial, para el cual Chile tiene en su mayoría el 10% del tiempo de observación (debido a que este método no puede ser utilizados en los observatorios que realizar sondeos del cielo, no es totalmente 10%).

2.4.1. ¿Qué necesita Chile para un OV?

En un estudio realizado el año 2008 [11], luego de un análisis de la situación nacional, se concluyeron un conjunto de líneas de acción para el establecimiento de un OV chileno que se describen a continuación.

2.4.2. Conexiones de redes académicas

Los desafíos planteados por la nueva instrumentación astronómica evidencian la necesidad de disponer de redes computacionales rápidas, que aseguren el acceso

óptimo a los observatorios, desde las instituciones usuarias. La tendencia mundial es a utilizar redes específicamente dedicadas a la investigación y las actividades académicas; estas redes, típicamente, proporcionan la conectividad troncal a nivel nacional, además de interconexiones a las redes académicas regionales.

La relevancia de la conexión de los grandes observatorios a las redes académicas se hace mucho más patente si se toma en consideración que tanto en Estados Unidos como en Europa los grandes observatorios han estado desde siempre conectados a las redes académicas.

2.4.3. *Grid*

La solución que a nivel mundial ha sido verificada como la más eficiente en términos de uso de recursos distribuidos para el desarrollo de múltiples actividades de investigación y procesamiento de datos, son las *Grid*.

Si bien, el procesamiento y almacenamiento de información no es el único objetivo de un *Grid*, es importante contar a lo menos con la infraestructura necesaria para el trabajo local. De esa forma, los recursos de almacenamiento actuarán como un caché mientras se realicen los procesamientos, pero no serán almacenados de forma local. Considerando la existencia de una infraestructura de red suficientemente capaz de sustentar estos procesos, sería posible utilizar los archivos de los observatorios para obtener los datos con los cuales trabajar localmente.

Lo anterior requiere de la definición de una infraestructura mínima por cada centro, de modo de garantizar su operación en colaboración con los demás recursos de la *Grid*. En un principio se puede utilizar aquella infraestructura ya existente, y generar o fortalecer alianzas con otros departamentos o programas fuera del área de la Astronomía, a fin de compartir estos recursos.

2.4.4. Creación de *Software* para OV

Las tareas relacionadas con un OV no se limitan al almacenamiento y procesamiento de datos. Si bien ellas son sumamente importantes, no pueden realizarse sin que exista toda una infraestructura de software que especifique la forma de llevar a cabo esos procesos. En ese sentido, es posible establecer alianzas con los proyectos de observatorio internacionales asentados en el país, a fin de proveerles los diversos tipos de software que se requieren para manipular y procesar los datos. Para ello es necesario participar activamente en el IVOA y aprovechar las actividades de formación que se dan en el marco de esta iniciativa. A partir de la adquisición de la experticia necesaria, es posible proveer, por ejemplo, pipelines de reducción de datos para los diferentes instrumentos.

2.4.5. Software para el Uso de la Grid

El uso de la tecnología *grid* requiere normalmente de desarrollos de software que permitan aprovechar de mejor modo los recursos ya existentes. En esto se pueden integrar iniciativas complementarias como CL-GRID y algunos proyectos de colaboración internacional.

2.4.6. Capital Humano

Para poder ser capaz de desarrollar todo lo anterior, es necesario gente experta en diversas áreas como astronomía, física, informática, etc. Para ello es necesario promover el asociamiento de estas disciplinas con la astronomía.

2.4.7. Incorporación al IVOA

Un punto importante es la incorporación de los astrónomos chilenos al IVOA. Para ello es necesario establecer algún tipo de institucionalidad para el OV nacional, a fin de que este sirva para representar a los astrónomos en las actividades del IVOA.

2.4.8. Roadmap

Se tiene un *roadmap* de la astronomía en Chile por Conicyt [12], que se detalla a continuación.

Promoviendo la astro-ingeniería en Chile

Chile puede aprovechar los tantos observatorios internacionales que se encuentran en su territorio, para fortalecer la educación en general, o la ciencia, tecnología, ingeniería y educación matemática. Chile puede actuar para estimular la astronomía chilena por sí misma. Desarrollo industrial puede ocurrir al capturar la innovación, mejoramiento de habilidades y desafíos ingenieriles dentro de la industria y expansión del servicio de las industrias que apoyan a los observatorios. Los observatorios internacionales nuevos y planificados en Chile, requieren una significativamente avanzada instrumentación científica

Se argumenta que la creación de una o dos iniciativas chilenas que lideren el desarrollo, entrega y uso de la instrumentación científica de nivel mundial de estos observatorios, tiene el potencial de mejorar significativamente el desarrollo de la tecnología en Chile. Adicionalmente a otras iniciativas que involucren educación, estímulo comercial, o creación de capacidades industriales de alta tecnología, se cree que un paso clave en mejorar la posición de Chile, se puede lograr manteniéndose cerca de las necesidades de los observatorios y posicionando a la comunidad astronómica chilena para que satisfaga dichas necesidades. Por esto, se debe entender que entendiendo las necesidades científicas y técnicas de los observatorios y sus comunidades (incluyendo a la comunidad chilena que tiene un tiempo de observación del 10%) y entregando instrumentación liderada por grupos chilenos, Chile puede proveer el apoyo para atraer, desarrollar y establecer uno o más grupos líderes en Astro-Ingeniería. Estos grupos podrán luego proponer y triunfar en una competencia revisada por sus pares, para desarrollar y entregar instrumentación. Esto puede llevar a la comunidad astronómica chilena a un mundo de ingeniería astronómica e

innovación. Éste será el primer paso para cualquier ambiente industrial que busque mejorar la astronomía en Chile. Para estimular la astronomía chilena no es necesario asociarse inmediatamente con observatorios internacionales o con la industria del sector, en un principio, el fortalecer los vínculos entre los estos actores es obviamente un deber y debe ser promovido.

La astronomía, además de utilizar matemática avanzada, y métodos analíticos y técnicos, requiere instrumentación especializada de última generación que puede tener aplicaciones en otras áreas como la minería, defensa nacional.

Pasos específicos en el *roadmap* para promover el desarrollo de tecnologías en el campo de la astronomía en Chile

- Incrementando o creando nuevas oportunidades de fondos a niveles más altos al de los fondos existentes en la actualidad.
- Realizando sondeos astronómicos por década de manera regular para así obtener un entendimiento claro de los intereses y las iniciativas programáticas de la comunidad local para el cercano, medio y largo plazo en el campo de la astronomía.
- Promover el desarrollo de instrumentaciones astronómicas indígenas.
- Atraer expertos internacionales a Chile.
 - Individuos cualificados con una carrera demostrada en el desarrollo de instrumentos y gestión de proyectos astronómicos de escala media/grande.
 - Atraer y facilitar la instalación de centros internacionales que tengan historial reconocido en el desarrollo de instrumentación astronómica.
- Promover el vínculo entre desarrolladores de tecnologías y sectores relevantes de la economía y la seguridad nacional.

Actualmente, la astronomía está mayoritariamente respaldada por Universidades, Colaboraciones Nacionales y Observatorios:

Universidades: Universidad de Chile, Pontificia Universidad Católica de Chile, Universidad de Concepción, Universidad de La Serena, Universidad Católica del Norte, Universidad de Valparaíso, Universidad Nacional Andrés Bello, Universidad Técnica Federico Santa María, Universidad de Tarapacá, entre otros.

Colaboraciones nacionales: SOCHIAS Sociedad Chilena de Astronomía, Centro Fondap de Astrofísica.

Observatorios: Observatorio Interamericano de Cerro Tololo, Observatorio Europeo Austral, Observatorio de la Institución Carnegie de Washington, Gemini, Southern Observatory for Astrophysical Research, Atacama Large Millimeter Array, Observatorio Chajnantor, Atacama Pathfinder Experiment, Atacama Submillimeter Telescope Experiment, Visible and Infrared Survey Telescope for Astronomy, Atacama Cosmology Telescope, Giant Magellan Telescope, Cornell Caltech Atacama Telescope, Large Synoptic Survey Telescope y Advanced Liquid-mirror Probe of Astrophysics, Cosmology and Asteroids.

2.4.9. Chilean Virtual Observatory

Durante los años 2013 - 2015 diversas instituciones chilenas en conjunto con ALMA, dirigidas por Mauricio Solar de la Universidad Técnica Federico Santa María son partícipes del proyecto FONDEF D11|1060-2013, el cual consiste en el “Desarrollo de una plataforma astroinformática para la administración y análisis inteligente de datos a gran escala”.

Las instituciones involucradas son Universidad de Chile, Pontificia Universidad Católica de Chile, Universidad de Concepción, Universidad Técnica Federico Santa María, Universidad de Santiago de Chile, ALMA y Consorcio Red Universitaria Nacional (REUNA).

La solución propuesta por el proyecto es “La solución propuesta será abordada principalmente realizando una investigación, análisis, diseño e implementación de alternativas de plataformas informáticas para un VO distribuido, que apoye a los astrónomos chilenos en la recolección y almacenamiento de sus datos en forma local, en el desarrollo de algoritmos eficientes de búsqueda semántica sobre el VO y el desarrollo de algoritmos de procesamiento inteligente de gran volumen de datos provenientes del VO.”

Los astrónomos e informáticos asociados al proyecto serán capaces de crear esta plataforma astroinformática. Donde esta memoria marca el inicio con el diseño conceptual de ChiVO.

2.5. ALMA

ALMA es un proyecto radioastronómico que consiste en 66 antenas a una altura de 5000 metros en el Desierto de Atacama, el cual tendrá Interferometría con líneas base de hasta 16 kilómetros, con un rango de onda desde los 3 milímetros hasta los 400 micrómetros (84 a 702 GHz). Es un proyecto multinacional construido por la ESO, NRAO y NAOJ en cooperación con Chile. Se espera que cuando esté totalmente operacional la cantidad de datos que generará será de: 6.6 MBytes/segundo = 200 TB/año desde la segunda mitad del 2013.

Actualmente, los datos generados son replicados en los ARCs (*ALMA Regional Centers*) y en el caso de los datos solicitados por los chilenos, éstos son obtenidos desde un ARC en el extranjero (USA, Japón o Europa) de vuelta a Chile, lo cual es un paso innecesario teniendo en cuenta que los datos son generados en Chile, por ello se realizará un OV para Chile en conjunto con ALMA.

Lo bueno es que el archivo científico de ALMA es relativamente nuevo y en ALMA pensaron en cómo se van a utilizar los datos, por lo que facilitan el uso de OV haciendo el archivo científico compatible con OV (ASA 2.1-R6).

Además, ALMA es parte del proyecto FONDEF D11|1060-2013, el cual consiste en el “Desarrollo de una plataforma astroinformática para la administración y análisis inteligente de datos a gran escala”. Esta memoria también es parte de este proyecto.

2.6. Servicios

Actualmente, en el mundo de la astronomía existen diversos servicios que facilitan el trabajo a los astrónomos. Para este proyecto, cabe destacar:

2.6.1. Resolvedores de Nombres

Simbad

Simbad [13] es una base de datos astronómica que provee datos básicos, identificaciones cruzadas, bibliografía y medidas de objetos astronómicos fuera del sistema solar.

El *Centre de Données astronomiques de Strasbourg* (CDS) ofrece buscar en Simbad a través de su herramienta resolvidora de nombres, Simbad recibe el nombre del objeto astronómico y éste retorna las coordenadas e información asociada, en caso de existir [14].

NED

El NASA/IPAC *Extragalactic Database* (NED) [15], es una base de datos multi-onda-largo de objetos extragalácticos provenientes de cientos sondeos del cielo y miles de publicaciones de investigación. Los datos que se proveen cubren completamente el espectro observable desde rayos gamma hasta radiofrecuencias.

A través de este servicio, es posible realizar búsquedas por nombre, objetos dentro de un radio del cielo alrededor del nombre de un objeto, objetos en un radio alrededor de una posición, coordenadas, parámetros, clasificaciones (tipos, atributos), código de referencia entre otras.

Es posible buscar en NED vía URL pasando los parámetros “RA”, “DEC” y “SR”, con los valores de las coordenadas RA, DEC y el radio de búsqueda. Todas las opciones de búsqueda y cómo realizarlas, se encuentran en el artículo realizado por Joseph Mazarella y el equipo de NED [16].

VizieR

VizieR [17] provee acceso a la librería más completa de catálogos astronómicos publicados y tablas de datos disponibles en línea (11575 catálogos). Es posible utilizar su servicio a través del CDS, utilizando el resolutor de nombres [14].

CDS Name Resolver

Herramienta del CDS que permite resolver nombres a través de un servicio web XML utilizando como fuentes Simbad, NED y/o VizieR [14]

CADC Name resolver

El *Canadian Astronomy Data Centre* (CADC) tiene una aplicación web que sirve peticiones HTTP y GET para resolver nombre de objetos astronómicos a coordenadas RA y DEC [20]. El CADC realiza las consultas a distintos servicios antes de entregar una respuesta y acepta varios parámetros. Los parámetros se describen a continuación.

Obligatorio:

target - Nombre del objeto a resolver.

Opcionales:

service - Servicio/s de los cuales se quiere resolver el nombre. Deben estar separados por coma si es más de uno. Por defecto, se consultan los tres servicios.

- *ned* - Consulta el servicio NED *service* en CalTech.

- *simbad* - Consulta el servicio SIMBAD en el CDS.
- *vizier* - Consulta los servicios de Vizier en el CADC y CDS.
- *all* - Consulta los servicios de NED, Simbad y Vizier.

format - Formato de los resultados de la consulta, puede ser ASCII (*ascii*) o XML (*xml*).

cached - Si son o no reutilizados resultados previos al realizar la consulta, por defecto se utilizan los resultados previos (*yes*), pudiendo escoger que no sea así (*no*)

LookUP

LookUp es un servicio web que entrega información de variadas fuentes como Simbad, imágenes de Flickr, World Wide Telescope de Microsoft, *papers* en el Astrophysics Data System o astro-ph y wikisky [34]. Dependiendo del formato de salida (Web, XML, JSON, AVM o *Mobile*) entrega información de distintas fuentes, pero siempre entregará las coordenadas en caso de encontrarse el objeto.

La forma de búsqueda se realiza vía URL, enviando los parámetros de el formato de salida con las opciones de: XML, JSON, AVM o Mobile, en el caso de Web, se debe dejar en blanco. Y el parámetro de entrada o “*name*” se le entrega al URL con el nombre del objeto deseado.

2.6.2. Acceso a bases de datos espectroscópicas

Splatalogue

El Splatalogue es un catálogo con cerca de 6 millones de líneas espectrales [21] que ofrece buscar vía web o a través de una interfaz compatible con Observatorios Virtuales llamada SLAP (*Spectral Line Access Protocol*, de IVOA) [22].

ADS

El SAO/NASA *Astrophysics Data System* (ADS) [23] es un portal de una librería digital para investigadores en Astronomía y Físicas, mantenido por el *Smithsonian Astrophysical Observatory* (SAO) subvencionado por la NASA. El ADS ofrece tres

bases de datos bibliográficas que contienen más de 10 millones de registros: Astronomía y Física, Física y arXiv e-prints.

ADS es un portal pero ofrece servicios de búsquedas a través de URL. Para ello se debe armar una URL con los parámetros necesarios. Una lista de los parámetros que se pueden utilizar se encuentra en el sitio del ADS [24] al igual que el manual de uso de este servicio [25].

ADS también ofrece servicio de API para desarrolladores que permite a aplicaciones acceder a los servicios de ADS, para ello es necesario registrarse en el portal de ADS y solicitar una llave de desarrollador, toda la información necesaria se encuentra en el repositorio Github de ADS [26].

2.7. Tipos de Archivos

2.7.1. FITS

El *Flexible Image Transport System* (FITS) [27] ha sido utilizado por los astrónomos por más de 30 años como formato para intercambiar, analizar y archivar datos digitalmente. Actualmente, se encuentra en su versión número 3 lanzada en Julio del 2008. Su estructura consta de 3 partes [28]:

1. Cabecera primaria con metadatos que describen la unidad de datos y la unidad de datos (HDU, siglas en inglés. Obligatorio).
2. Extensiones (Opcional), cada extensión tiene su propia cabecera y datos.
3. Otros registros especiales (Opcional, restringido).

2.7.2. VOTable

El formato VOTable es un estándar XML para la transferencia de datos representados como un conjunto de tablas. Básicamente es un conjunto de filas, cada una con una estructura especificada en la tabla de metadatos. Cada fila está

compuesta de una secuencia de celdas que contienen o un tipo de datos primitivo o un arreglo de éstos. VOTable fue diseñado teniendo siempre en mente el formato FITS. La estructura sería de esta manera:

Tabla 2.3. Estructura de un VOTable.

Tabla	Lista de campos + Tabla de datos
Tabla de datos	Filas
Fila	Lista de celdas
Celda	Primitivo
	O lista de primitivos de largo variable
	O arreglo multidimensional de primitivos
Primitivo	Entero, carácter, flotante, etc.

2.7.3. MS

Formato enfocado en tipo de datos radioastronómicos tanto *single-dish* como de interferometría. Es capaz de soportar cualquier tipo de radiotelescopio. Básicamente es un sistema de base de datos relacional con una estructura predefinida de una tabla principal con 12 sub-tablas obligatorias y otras opcionales. Su desarrollo se encuentra estable pero inactivo.

2.7.4. ASDM

El ALMA *Science Data Model* (ASDM), a diferencia del MS, es específico para datos provenientes de ALMA. Este formato contiene toda la información obtenida durante una observación que podría ser necesaria para un análisis científico. Se puede convertir a MS a través de CASA. Incluye tres tipos de datos importantes:

1) Correlador/*backend*: datos observacionales

e.j., *raw visibilities, total power readings, etc. (the bulk of the data)*.

2) Metadatos: datos necesarios para describir el proceso de observación

e.j., *source information, spectral setup*.

3) Auxiliares: Data de monitoreo obtenida durante la observación y necesitada para la reducción

e.j., clima, *pointing information*.

Su desarrollo se encuentra estable pero activo.

Para entender qué es lo que contiene un ASDM, es necesario entender cómo están encapsulados los datos generados por ALMA y el flujo desde que se genera una propuesta hasta que se realiza la observación, lo cual se explica en la Figura 2.4.

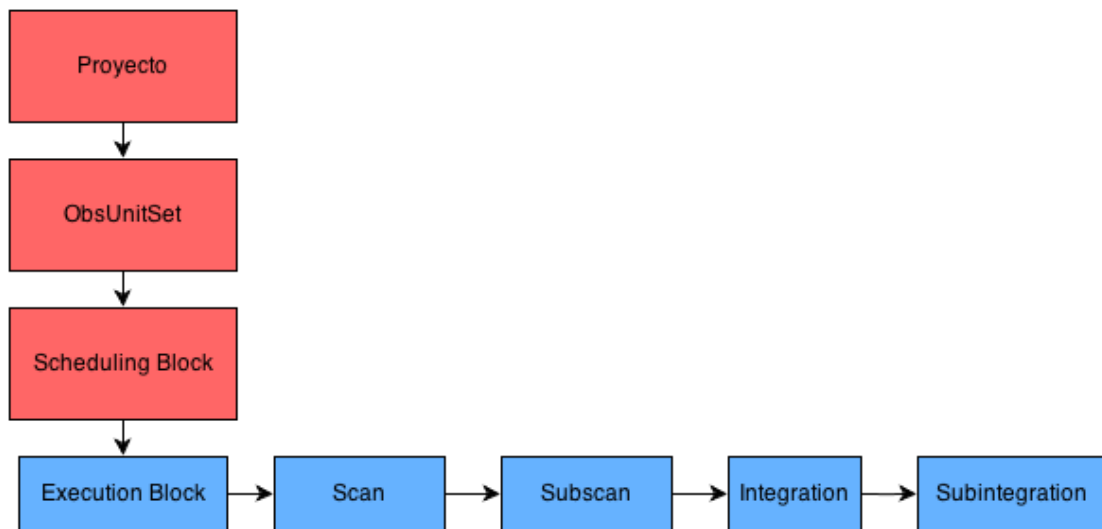


Figura 2.4. Jerarquía de los “objetos observados” usados en el ASDM.

En la Figura 2.4. cada nivel debe incluir al menos uno del nivel inferior. Los cuadrados rojos son objetos creados antes del tiempo de observación y los azules después.

Un astrónomo envía a ALMA un **proyecto**, que debe ser dividido en uno o más bloques llamados **bloques de programa**, a su vez, éstos últimos son divididos en uno o más **bloques de planificación**, ALMA organiza estos bloques de planificación acorde a diversos factores como la posición de las antenas y la calidad del clima,

entre otros. Los bloques de planificación, al momento de ser ejecutados, se convierten en **bloques de ejecución** que contiene datos asociados a la observación. Entonces, un ASDM contiene datos de observación asociados a un bloque de ejecución o *execution block*. Además, un ASDM también contiene uno más **scan**, un scan tiene uno o más **subscan**, de igual manera con las **integraciones** y **subintegraciones**.

Capítulo 3: Diseño

3.1. Necesidades y Requerimientos

Luego de múltiples entrevistas, charlas y un *workshop*, se definen las siguientes necesidades por parte de los astrónomos chilenos en general.

3.1.1. Necesidades

Descubrir: Encontrar datos astronómicos de un objeto y/o instrumento sobre una región específica del espacio de alta dimensión, en base a parámetros de los ejes espaciales, temporales, espectrales, corrimiento al rojo, polarización, etc. Ya sea por búsqueda o por exploración.

Obtener: Enlace a descarga de los datos requeridos. Ya sea en ChiVO o en un servicio externo.

Comparar: Datos obtenidos entre distintas fuentes de información.

Analizar: Minería de datos, etc.

3.1.2. Requerimientos

Con los cuatro puntos anteriores como base, en conjunto con el alcance de esta memoria, se describen los siguientes requerimientos que se encuentran evaluados en:

Necesidad: Esencial, deseable, opcional.

Prioridad temporal: Alta, media, baja.

1. Buscar por coordenadas o región del cielo [Necesidad: Esencial | Prioridad temporal: Alta]:

Se podrán realizar búsquedas de posición mediante coordenadas y radio angular (cónicas) o por región del cielo.

Los parámetros de las coordenadas pueden ser en distintos sistemas como ecuatorial, eclíptico, galáctico o supergaláctico. Los parámetros ingresados se convertirán al sistema de la fuente de datos, para así poder realizar las búsquedas, IVOA utiliza los sistemas de coordenadas ICRS y ecuatoriales J2000.

En un principio, el sistema ofrecerá servicio de búsqueda por coordenadas cónicas y más adelante, en caso de ser vía un portal web, se ofrecerá búsqueda por región de cielo.

El sistema también deberá permitir buscar simultáneamente un listado de coordenadas en un formato predeterminado.

2. Buscar por nombre o tipo de objeto [Necesidad: Esencial | Prioridad temporal: Alta]:

El sistema deberá permitir buscar por nombres de objetos que se encuentren definidos en **Sesame**, del *Centre de Données astronomiques de Strasbourg* (CDS).

Por otro lado, la búsqueda por tipo o subtipo de objeto, tales como estrellas en formación, estrellas nebulosas planetarias, supernovas, galaxias, cometas, entre otros, permitirá al usuario encontrar datos relacionados con una problemática en especial. En un principio, el sistema realizará estas búsquedas acorde a la información presente en los catálogos.

A futuro, el sistema deberá permitir minería de datos para detección de tipos de objetos similares, esto es posible debido a que existen clasificaciones discretas que permiten clasificar los objetos que se encuentran en las observaciones.

Este tipo de búsquedas, se transforman en búsquedas por coordenadas, ya que al buscar por un nombre, por ejemplo, Sesame responde con la correspondiente ubicación del objeto en coordenadas. Y luego de ello se procede a realizar la búsqueda por coordenadas correspondiente.

El resultado de la búsqueda deberá facilitar la obtención de datos para ser analizados como secuencias de tiempo.

3. Buscar por metadatos espectrales (frecuencia y resolución) [Necesidad: Esencial | Prioridad temporal: Alta]:

Se podrán realizar búsquedas por metadatos espectrales, lo cual consiste en búsqueda por banda o rango de frecuencia, búsquedas por líneas espectrales y corrimiento al rojo o búsquedas por resolución espectral. Específicamente, hay dos enfoques:

Galáctico: Por frecuencia en reposo y velocidad radial.

Extragaláctico: Por frecuencia en reposo y corrimiento al rojo.

La Frecuencia en reposo incluye búsquedas por molécula, transición de molécula (vibracional, rotacional o electrónica) o frecuencia de línea espectral.

4. Buscar por metadatos espaciales (resolución angular y campos de visión) [Necesidad: Esencial | Prioridad temporal: Alta]:

Se podrán realizar búsquedas espaciales en base a parámetros relacionados con rangos de resolución angular y campos de visión y siempre en base a coordenadas.

Los parámetros de las coordenadas pueden ser en distintos sistemas como ecuatorial, eclíptico, galáctico o supergaláctico. Los parámetros ingresados se convertirán al sistema de la fuente de datos, para así poder realizar las búsquedas.

Además, se podrá especificar parámetros relacionados con la forma de las observaciones (rectangulares o redondas).

5. Buscar por metadatos temporales [Necesidad: Esencial | Prioridad temporal: Baja]:

Se podrán realizar búsquedas por metadatos temporales que pueden ser clasificadas en dos tipos de búsquedas:

5.1. Cuando fue realizada la observación, incluyendo cuantas veces se observó un objeto y/o el intervalo entre observaciones.

5.2. Nivel de ruido, duración de la observación o tiempo de integración. Dado un ruido, se necesita un tiempo de integración que depende de la frecuencia observada y del clima.

6. Buscar por polarización [Necesidad: Esencial | Prioridad temporal: Media]:

Cada imagen se puede dividir en cuatro parámetros llamados los parámetros de Stokes o en dos: izquierda y derecha. En radioastronomía no suele hacerse debido a que requiere una alta precisión del instrumento, en el caso de ALMA se requiere que esté lista la calibración.

El sistema debe permitir buscar si existe o no polarización en alguno de los parámetros de Stokes: I, Q, U o V.

7. Cruzamiento de información [Necesidad: Esencial | Prioridad temporal: Alta]:

La búsqueda cruzada debe permitir al usuario realizar las búsquedas mencionadas anteriormente en múltiples fuentes de datos distribuidos globalmente, sin importar su tipo. Lo que permitirá obtener todos los datos existentes sobre un objeto o área espacial y así evitar realizar observaciones innecesarias debido al descubrimiento de observaciones existentes. Los tipos de fuentes pueden ser Sesame, ALMA u Observatorios Virtuales, que cumplan con los estándares de IVOA.

La búsqueda de un mismo objeto en distintas fuentes de información, debido a que en cada fuente el instrumento tiene un margen de error en cuanto a la posición del objeto, debe ser capaz de realizar una intersección entre los radios de margen de error de las distintas fuentes para identificar al objeto en una búsqueda cruzada.

8. Simulaciones [Necesidad: Deseable | Prioridad temporal: Baja]:

Es a veces necesario realizar comparaciones con observaciones obtenidas a través de simulaciones, así como es costoso realizar una observación dos veces, para una simulación puede ser aún más, debido a que dependiendo de la magnitud, hay simulaciones que requieren una gran capacidad de cómputo.

9. Servicios Bibliográficos [Necesidad: Deseable | Prioridad temporal: Baja]:

Las herramientas existentes cumplen su función correctamente. Por ello, basta con que al buscar un objeto, se desplieguen también resultados de investigaciones que se hayan realizado al respecto con un enlace a SIMBAD o similares.

3.2. Casos de Uso

Luego de haber identificado las necesidades y requerimientos del ChiVO en conjunto con astrónomos chilenos expertos en el área, se procede a identificar los casos de uso asociados a cada requerimiento. Para ello se realizan múltiples entrevistas con dichos expertos y luego se procede a buscar validación de los casos de uso identificados con los distintos actores identificados.

3.2.1. Actores

Los actores identificados están clasificados en dos funcionalidades: Consumir datos y proveer datos. A partir de ello surge el actor principal consumidor, del cual heredan los actores astrónomo profesional y amateur, ambos actores realizan búsquedas sobre ChiVO pero con objetivos distintos. En el caso de los actores proveedores, dos heredan del actor proveedor, en un principio ALMA (a futuro se podrán integrar más proveedores instrumentales) y el actor OV-compatible, pudiendo ser otro OV o un servicio que sea compatible con el ecosistema de OV.

3.2.2. Casos de uso

Para simplificar los casos de uso, en la Figura 3.1, los casos de uso se han dividido en categorías acorde a los requerimientos establecidos en la sección anterior y aquellos casos de uso que son transversales a éstos.

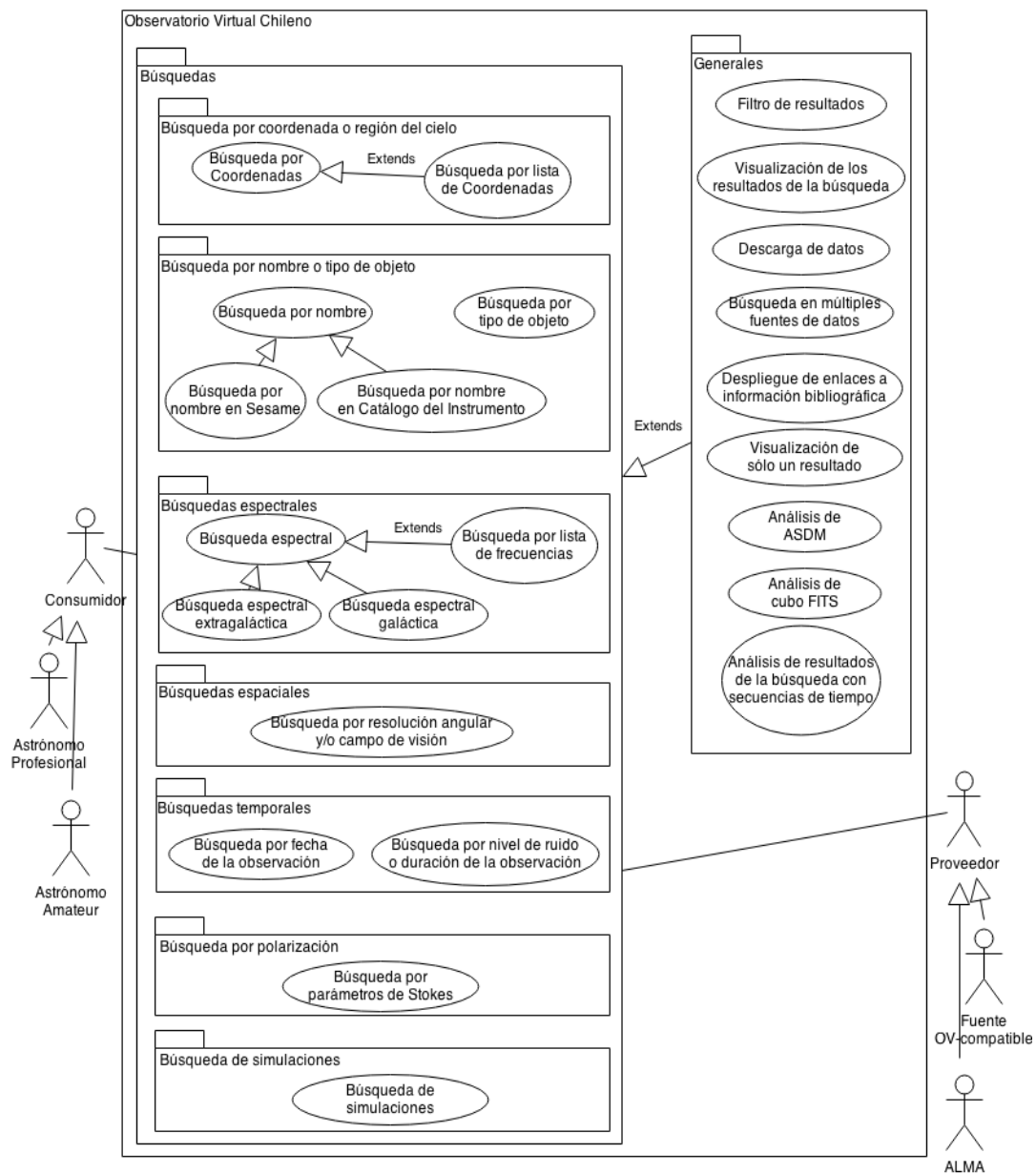


Figura 3.1. UML de Casos de Uso para el ChiVO.

Como se muestra en la Figura 3.1, los casos de uso transversales corresponden al paquete Generales y el resto que corresponden a los requerimientos antes expuestos, se encuentran clasificados dentro del paquete Búsquedas. También se evidencia la herencia de actores y cómo interactúan con los casos de uso. Los casos de uso son descritos en mayor detalle en la Tabla 3.1 y la Tabla 3.2 con columnas con el nombre

del caso de uso, requerimiento o paquete al que hace referencia, descripción, necesidad y prioridad.

Tabla 3.1. Casos de uso transversales a los requerimientos planteados evaluados por necesidad y prioridad.

ID	Casos de Uso (aplican a casi todos)	Req. Ref.	Descripción	Necesidad	Prioridad
1.1	El usuario filtra los resultados de la búsqueda en el Portal Web	Del 1 al 7	Ya que el usuario puede recibir una gran cantidad de resultados, éste debe ser capaz de realizar un filtro sobre alguna columna de los datos recibidos al realizar la búsqueda para así poder identificar los datos que se encuentre buscando. Este filtro puede ser realizado en el mismo navegador del usuario o a través de una nueva consulta.	Esencial	Alta
1.2	El usuario descarga datos desde el Portal Web	Del 1 al 7	Una vez que el usuario encuentre los datos que requiere, éste procede a descargarlos. Para ello el sistema provee un enlace directo a la fuente de los datos.	Esencial	Alta
1.3	El usuario visualiza una representación gráfica que compara los metadatos de los resultados de la búsqueda	Del 1 al 7	Una vez que el usuario reciba un conjunto de resultados, podrá visualizarlos en gráficos acordes al tipo de búsqueda y en base a los metadatos recibidos por la búsqueda.	Esencial	Media
1.4	El usuario visualiza un producto o subproducto de los datos de la observación presente en los metadatos	Del 1 al 7	Visualización de la forma geométrica de las observaciones (rectangulares o redondas); cobertura UV; Calibración de paso de banda; Espectro observado; Imagen observada en un plano del producto cuando sea posible (podría ser parte de análisis en vez de observación);	Esencial	Alta
1.5	Análisis de cubo FITS	Del 1 al 7	Selecciona píxeles; proyección en una dimensión; flujo en un área; series de tiempo.	Esencial	Alta
1.6	Análisis de ASDM	Del 1 al 7	Selecciona voxels; proyección en una dimensión; flujo en un área; series de tiempo.	Deseable	Media
1.7	Los resultados de la búsqueda deben ser analizables para secuencias de tiempo	Del 1 al 7	El usuario luego de realizar una búsqueda por tipo y/o subtipo de objeto, puede obtener los resultados de la búsqueda para ser analizados como secuencias de tiempo.	Esencial	Baja
1.8	El usuario busca un objeto en múltiples fuentes de datos como archivos de los Satélites Spitzer y Herschel	7	El usuario ingresa al Portal Web, rellena los campos de nombre o código de objeto o coordenadas junto con un radio y se realiza la búsqueda. El sistema deberá realizar la búsqueda en base a ello en múltiples fuentes de información. Si se ingresa el nombre o código, se realiza una búsqueda teniendo en cuenta el margen error de cada fuente de datos.	Esencial	Media
1.9	El usuario al realizar una búsqueda, se le despliega en los resultados enlaces a fuentes de información bibliográfica.	9	El usuario ingresa al Portal Web y al desplegarse los resultados de una búsqueda, cada resultado contiene un enlace a una búsqueda con todos los datos contenidos en SIMBAD, ADS y/o NED pertenecientes al/los objeto/s.	Deseable	Media

En la Tabla 3.1 se detallan los casos de uso que extienden al resto, como se puede apreciar en la Figura 3.1 a través de la relación *extends*, debido a que son casos de uso que describen la manipulación de los resultados arrojados por los casos de uso de búsquedas. La mayoría de estos casos de uso son esenciales y de prioridad alta debido a que forman parte de las funcionalidades básicas de un OV para un mejor manejo sobre los resultados de las búsquedas y así lograr que el usuario encuentre lo que necesite con mayor facilidad, esto es acorde a las entrevistas realizadas con los radioastrónomos y usuarios de OV.

En la Tabla 3.2 y Tabla 3.3 desplegadas a continuación, se encuentran los casos de uso pertenecientes al paquete de búsquedas de la Figura 3.1. Éstos corresponden a distintos tipos de búsqueda con parámetros con distintos enfoques físicos como el espacio, tiempo o frecuencia. Éstos se encuentran ordenados en importancia decreciente acorde a la necesidad y prioridad definida por los radioastrónomos, en la Tabla 3.2 se encuentran los más relevantes y en la Tabla 3.3 los menos relevantes acorde a los requerimientos asociados. Además, como se puede apreciar en la Figura 3.1, estos casos de uso son extendidos por los casos de uso generales, ya que los casos de uso generales trabajan con los resultados arrojados por estas búsquedas.

Los casos de uso en general, representan distintos tipos de búsquedas que pueden ser realizadas simultáneamente, por ejemplo, en un formulario de búsqueda en el sitio web de ChiVO, donde se tienen distintas cajas de *input* de texto, se podrían ingresar parámetros de coordenadas espaciales y de un rango de frecuencia al mismo tiempo, a pesar de que están descritos como casos de uso distintos asociados a distintos requerimientos.

Tabla 3.2. Lista de casos de uso más relevantes asociados a cada requerimiento evaluados por necesidad y prioridad.

ID	Nombre del Requerimiento / Caso de Uso	Req. Ref.	Descripción	Necesidad	Prioridad
2	Buscar por coordenadas o región del cielo	-			
2.1	El usuario ingresa al Portal Web y realiza una búsqueda por coordenadas	1	El usuario ingresa al Portal Web, rellena los campos de coordenadas y radio angular (búsqueda cónica) o región de cielo y realiza una búsqueda. Los parámetros de coordenada pueden pertenecer al sistema ecuatorial (J2000 o B1950), eclíptico, galáctico o supergaláctico.	Esencial	Alta
2.2	El usuario realiza una búsqueda de listado de coordenadas	1	El usuario ingresa al Portal Web, ingresa una lista de coordenadas y radios angulares o regiones de cielo y realiza una búsqueda. Los parámetros de coordenada pueden pertenecer al sistema ecuatorial (J2000 o B1950), eclíptico, galáctico o supergaláctico. También el usuario puede subir un archivo con un formato establecido por el sitio con el listado de coordenadas.	Esencial	Alta
3	Buscar por nombre o tipo de objeto	-			
3.1	El usuario ingresa al Portal Web y realiza una búsqueda por nombre en base a Sesame	2	El usuario ingresa al Portal Web, rellena el campo de nombre según los nombres definidos en Sesame y realiza una búsqueda.	Esencial	Alta
3.2	El usuario ingresa al Portal Web y realiza una búsqueda por nombre en base a catálogo de un instrumento	2	El usuario ingresa al Portal Web, rellena el campo de nombre según los nombres definidos en Catálogos específicos de un instrumento y realiza una búsqueda.	Esencial	Alta
3.3	El usuario ingresa al Portal Web y realiza una búsqueda por tipo de objeto en un área del cielo	2	El usuario ingresa al Portal Web, rellena los campos de tipo y/o subtipos de objetos y un área del cielo y realiza una búsqueda.	Esencial	Alta
4	Buscar por metadatos espectrales	-			
4.1	El usuario ingresa al Portal Web y realiza una búsqueda espectral extragaláctica	3	El usuario ingresa al Portal Web, rellena los campos de banda o rango de frecuencia; y/o líneas espectrales y corrimiento al rojo (z); y/o resolución espectral y/o ruido y realiza una búsqueda.	Esencial	Alta
4.2	El usuario ingresa al Portal Web y realiza una búsqueda espectral galáctica	3	El usuario ingresa al Portal Web, rellena los campos de banda o rango de frecuencia; y/o líneas espectrales y velocidad radial (v_r); y/o resolución espectral y/o ruido y realiza una búsqueda. Línea espectral incluye campos de moléculas, transición de moléculas (vibracional, rotacional o electrónica) o frecuencia en reposo de la línea espectral.	Esencial	Alta
4.3	El usuario realiza una búsqueda de listado de frecuencias	3	El usuario ingresa al Portal Web, ingresa una lista de frecuencias y realiza una búsqueda. También el usuario puede subir un archivo con un formato establecido por el sitio con el listado de frecuencias.	Esencial	Alta

Tabla 3.3. Lista de casos de uso asociados a cada requerimiento evaluados por necesidad y prioridad.

ID	Nombre del Requerimiento/ Caso de Uso	Req. Ref.	Descripción	Necesidad	Prioridad
5	Buscar por metadatos espaciales (resolución angular y campos de visión)	-			
5.1	El usuario ingresa al Portal Web y realiza una búsqueda por resolución angular y/o campos de visión	4	El usuario ingresa al Portal Web, rellena los campos de resolución angular y/o campos de visión y realiza una búsqueda.	Esencial	Medio
6	Buscar por metadatos temporales	-			
6.1	El usuario ingresa al Portal Web y realiza una búsqueda relacionada con cuando fue realizada la observación	5	El usuario ingresa al Portal Web, rellena los campos de tiempo, cantidad de observaciones de un objeto y/o el intervalo entre las observaciones y realiza la búsqueda.	Deseable	Baja
6.2	El usuario ingresa al Portal Web y realiza una búsqueda relacionada con nivel de ruido, duración de la observación o tiempo de integración.	5	El usuario ingresa al Portal Web, rellena los campos de nivel de ruido, duración de la observación o tiempo de integración y realiza una búsqueda.	Esencial	Alta
7	Buscar por polarización	-			
7.1	El usuario ingresa al Portal Web y realiza una búsqueda por parámetros de Stokes	6	El usuario ingresa al Portal Web, rellena los campos de parámetros de Stokes necesarios (I, Q, U y/o V) o de polarización izquierda o derecha y realiza la búsqueda.	Esencial	Baja
8	Simulaciones	-			
8.1	El usuario realiza una búsqueda de una simulación	8	El usuario ingresa al Portal Web y realiza un búsqueda de una simulación.	Deseable	Baja

El paso siguiente de la metodología corresponde a definir un modelo de datos que sea capaz de representar los elementos que deban ser representados y las relaciones entre ellos. Para ello, es necesario definir un modelo de datos que cumpla con lo requerido en los casos de uso, éste debe contener todos los campos necesitados en los requerimientos y sus casos de uso. Sin embargo, los datos a almacenar en este modelo deben existir en los datos de ALMA en sus archivos ASDM, para así poder llenar el modelo de datos de ChiVO, de otra forma sería imposible y requeriría una validación adicional de los casos de uso con los radioastrónomos, ya que sería necesario realizar cálculos que podrían depender de cada usuario. En la Tabla 3.4 se despliega dicha información.

Tabla 3.4. Lista de campos asociados a los requerimientos y ubicación de éstos dentro del ASDM o servicios astronómicos.

Requerimiento	Campo	ASDM
Coord	coordenadas	Source.direction
Coord	radio angular	Formulario
Coord	región del cielo	Formulario
Nombre o tipo	nombre objeto Sesame	Sesame
Nombre o tipo	nombre objeto ALMA	Source.sourceName
Nombre o tipo	Tipo objeto	Simbad
Espectral	Banda	receiver.frequencyBand
Espectral	Rango Frecuencia Observado	spectralWindow.refFreq - spectralWindow.totBandwidth/2 , spectralWindow.refFreq + spectralWindow.totBandwidth/2
Espectral	Líneas espectrales	spectralwindow.linearray
Espectral	Corrimiento al rojo	Diego preguntará.
Espectral	Resolución espectral	spectralWindow.Resolution
Espectral	Velocidad radial	source.Sys_Vel
Espectral	Nombre Molécula	Splatalog
Espectral	Transición vibracional molécula	Splatalog
Espectral	Transición rotacional molécula	Splatalog
Espectral	Transición electrónica molécula	Splatalog
Espectral	frecuencia en reposo de la línea espectral	rango entre source.restFrequency[0] hasta el source.restFrequency[numLines]
Espacial	Resolución angular	$1.2 * \lambda / (\text{separación máxima entre las antenas})$
Espacial	Campos de visión	$1.2 * \lambda / \text{Diámetro de una antena}$
Temporal	Tiempo de la observación	main.time
Temporal	Cantidad de observaciones	Resultados
Temporal	Nivel de Ruido	source.fluxErr
Temporal	Duración de la observación en la fuente (on source)	Main.interval * Main.numIntegration
Polarización	Parámetros de Stokes (I,Q,U,V)	Source.stokesParameter[numStokes]
Polarización	Polarización izquierda	polarizationType
Polarización	Polarización derecha	polarizationType

En la Tabla 3.4 se muestra la lista de requerimientos, los campos asociados a cada requerimiento y la ubicación dentro del ASDM, el nombre de algún servicio astronómico externo, parámetro de filtro u orden en el despliegue de resultados o parámetros de búsqueda. Esta lista fue confeccionada en conjunto con Diego

Mardones (radioastrónomo chileno) y Kurt Plarre (ALMA). Se concluye que todos los parámetros de búsqueda requeridos por la comunidad radioastronómica para realizar búsquedas sobre datos de ALMA se encuentran en los metadatos de los ASDM o son fácilmente obtenibles mediante cálculos sencillos.

3.3. Diseño de modelo de datos

Un modelo de datos es una descripción de un conjunto de entidades que se deben almacenar para un uso determinado [29], y en él se detallan los datos que serán almacenados y la relación entre éstos de manera estructurada. Es el pilar más fundamental de un OV ya que según West y Fowler “si está hecho de manera consistente a lo largo de los sistemas, entonces la compatibilidad entre los datos podrá ser lograda. Si la misma estructura de datos son usadas para almacenar y acceder los datos entonces, distintas aplicaciones pueden acceder a éstos.”

Para definir los modelos de datos, se analizaron las necesidades de búsqueda presentadas anteriormente a través de los requerimientos, casos de uso y finalmente a través de la Tabla 3.4 con la lista de campos asociados a cada requerimiento. A continuación, se realizó una investigación de los estándares de IVOA y los tipos de búsquedas que cubre cada uno. Así, los estándares de modelo de datos *Observation Core* y *Characterisation* fueron escogidos ya que sus enfoques principales son el descubrimiento y análisis/procesamiento de datos. Además, estos estándares son lo suficientemente generales como para aplicarlos a casi cualquier tipo de observación astronómica, por lo que es una buena base para ChiVO.

En la Figura 3.2, se muestra un diagrama de alto nivel de la relación entre las clases.

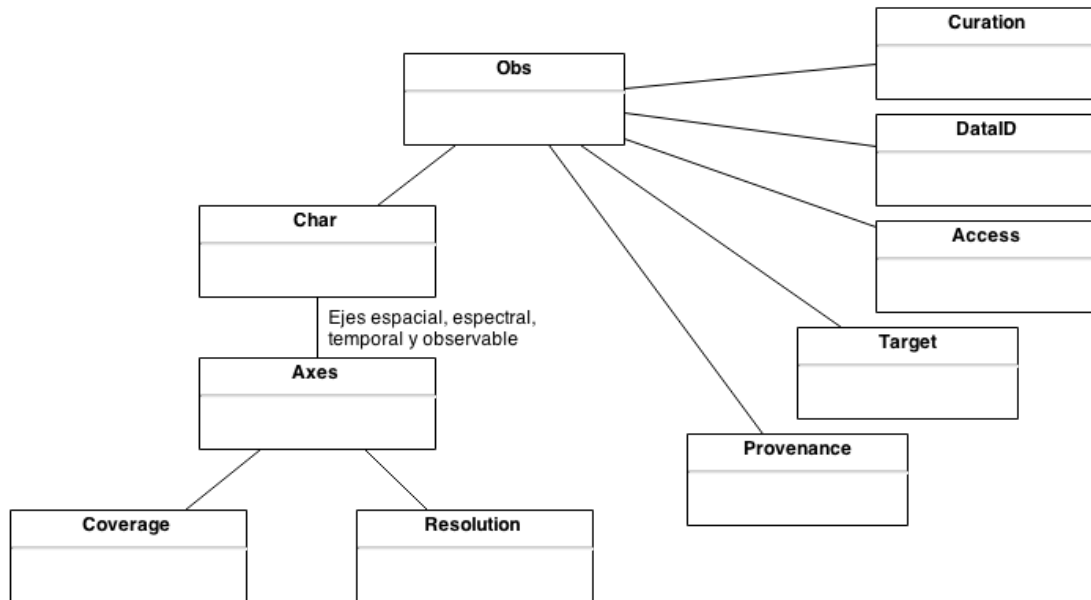


Figura 3.2. Clases principales del modelo de datos de ChiVO.

La estructura de la Figura 3.2 en conjunto con otras tablas ya han sido utilizadas en el área científica, por ejemplo para el RADAMS [30] (DSS63 y *Robledo Archive Data Model*). Las clases *Obs*, *Curation*, *DataID*, *Access*, *Target* y *Provenance* corresponden al estándar de IVOA *Observation Core DM*, lo cual permite comparar con casi cualquier tipo de datos astronómico. Las clases *Char*, *Axes*, *Coverage* y *Resolution* corresponden al estándar de IVOA *Characterisation DM* y permite realizar búsquedas más especializadas, ya que describen más detalladamente los datos, esto también permite realizar análisis a gran escala sobre los metadatos. Cabe mencionar que estos datos pueden representar, en el caso de ALMA, a los ASDM o a los archivos FITS.

Para poder definir en detalle cada tabla del modelo, hay que tener en cuenta que cada estándar de modelo de datos del IVOA propone tres niveles de requisitos y que al cumplirlos, se es compatible con el ecosistema OV en distintos grados: si se cumplen todos los requisitos obligatorios, se es compatible; todos los recomendados, se es totalmente compatible; finalmente los opcionales no afectan.

El acceso a los datos de ChiVO puede ser a través de TAP (*Table Access Protocol*)[31] o SSAP (*Simple Spectral Access Protocol*) [32].

3.3.1. Observación

El conjunto de clases de la observación corresponde a la representación del estándar de modelo de datos *Observation Core* de IVOA, y es la más importante del modelo de datos de ChiVO ya que contiene lo básico de toda observación y aplica para cualquier tipo de datos, no sólo provenientes de ALMA o radioastronomía, esto permite descubrimientos nunca antes realizados y también búsquedas que abarcan una gran cantidad de observaciones, independiente del instrumento que generó los datos. Es el siguiente paso de la astronomía.

Para el caso de ALMA, o cualquier instrumento, este conjunto de clases representa una observación pero surge un problema al intentar definir una observación. En ALMA, desde que el científico envía una propuesta de científica hasta que recibe se adquieren los datos, se pasan por distintas etapas, las cuales fueron descritas en la Figura 2.4 en el Estado del Arte.

Se puede definir una observación en base a los productos que ALMA entrega a los astrónomos y en este caso, una observación podría ser respecto a un ASDM o a un FITS el uso del uno o el otro dependerá del enfoque que ChiVO tome ya sea por razones científicas o técnicas. En el caso del ASDM, este contiene un bloque de ejecución (*Execution Block*) y en el caso del FITS, actualmente ALMA entrega un Scan asociado a una ventana espectral (rango de frecuencia observada), que debido a las capacidades actuales de alma son cuatro.

Sin importar el enfoque que tome ChiVO, el estándar de modelos de datos IVOA *Observation Core* requiere ciertos campos para que la implementación de éste sea

compatible con el ecosistema de los observatorios virtuales, es por ello que se presentan ambas opciones en este trabajo. En la Tabla 3.5 se describen las columnas obligatorias del estándar de modelo de datos *Observation Core* de IVOA.

Tabla 3.5. Columnas obligatorias del estándar de modelo de datos *Observation Core* de IVOA para ser compatible con el ecosistema OV.

Nombre de la Columna	Descripción	Clase
dataprodukt_type	Tipo de producto de los datos lógicos (imagen, cube, etc.)	Obs
calib_level	Nivel de calibración {0, 1, 2, 3}	Obs
obs_collection	Nombre de la colección de datos	DataID
obs_id	ID de la observación	DataID
obs_publisher_did	Identificador del conjunto de datos dado por el que publica	Curation
access_url	URL para acceder (descargar) el conjunto de datos	Access
access_format	Formato del contenido del archivo	Access
access_estsize	Tamaño estimado del conjunto de datos en kilo bytes	Access
target_name	Objeto astronómico observado, en caso de que haya	Target
s_ra	Ascensión central derecha, ICRS	Char.SpatialAxis.Coverage.Location.coord.value
s_dec	Declinación central, ICRS	Char.SpatialAxis.Coverage.Location.coord.value
s_fov	Diámetro (límites) de la región cubierta	Obs
s_region	Región cubierta como es especificado en STC o ADQL	Char.SpatialAxis.Coverage.Support.Area
s_resolution	Resolución espacial de los datos como FWHM	Char.SpatialAxis.Coverage.Resolution.refVal
t_min	Tiempo de inicio en MJD	Char.TemporalAxis.Coverage.Bounds.coord.minValue
t_max	Tiempo de fin en MJD	Char.TemporalAxis.Coverage.Bounds.coord.maxValue
t_exptime	Tiempo total de exposición	Char.TemporalAxis.Coverage.Support
t_resolution	FWHM de la resolución temporal	Char.TemporalAxis.Resolution.refVal
em_min	Inicio en coordenadas espectrales	Char.SpectralAxis.Coverage.Bounds.coord.minValue
em_max	Fin en coordenadas espectrales	Char.SpectralAxis.Coverage.Bounds.coord.maxValue
em_res_power	Poder de resolución espectral	Char.SpectralAxis.Resolution.refVal
o_ucd	UCD de lo observable (ej. phot.flux.density)	Char.ObservableAxis.ucd
pol_states	Lista de los estados de la polarización (NULL si no aplica)	Obs
facility_name	Nombre de la instalación utilizada para esta observación	Provenance
instrument_name	Nombre del instrumento utilizado para esta observación	Provenance

En la Tabla 3.5 se indican junto a una breve descripción los campos que son requeridos por el estándar. En la primera columna de la tabla se encuentra el nombre de la campo, en la segunda columna una breve descripción y en la tercera donde se puede encontrar el campo, ya sea dentro del modelo de datos de observación o de caracterización. Cabe notar que estás son las columnas mínimas requeridas y hay otras que son sugeridas, agregar estas columnas no supone un efecto mayor sobre lo presentado en este trabajo, estas columnas se encuentran descritas en el estándar de IVOA [33].

Cada dato contenido en los campos mencionados anteriormente deben ser obtenidos de alguna manera a partir de los datos de ALMA, es por ello que es necesario investigar en que parte del ASDM o FITS se encontrarán. Dicha investigación fue realizada y su resultado se describe en la Tabla 3.6.

Tabla 3.6. Campos del estándar ObsCoreDM y origen en el caso de ASDM y FITS.

Nombre de la Columna	ALMA (ASDM)	FITS
dataproduct_type	visibility	cube
calib_level	1	1
obs_collection	ALMA	ALMA
obs_id	[ExecBlock.execBlockUID]	[ExecBlock.execBlockUID]
obs_publisher_did	[Cycle ID]	[Cycle ID]
access_url	[URL de ChiVO]	[URL de ChiVO]
access_format	application/x-asdm	application/fits
access_estsize	[main.dataSize]	[[tamaño del archivo]]
target_name	[Source.sourceName]	[Source.sourceName]
s_ra	[Source.direction]	[Source.direction]
s_dec	[Source.direction]	[Source.direction]
s_fov	[1.2 * lambda / Diámetro antena]	[1.2 * lambda / Diámetro antena]
s_region	circle	circle
s_resolution	[1.2*lambda/(ExecBlock.baseRangeMax)]	[1.2*lambda/(ExecBlock.baseRangeMax)]
t_min	[ExecBlock.startTime]	[[tiempo de inicio]]
t_max	[ExecBlock.endTime]	[[tiempo final]]
t_exptime	[main.interval]	[[tiempo de inicio - tiempo final]]
t_resolution	[mainTable.interval]	[mainTable.interval]
em_min	[ExecBlock.baseRangeMin]	[[menor frecuencia]]
em_max	[ExecBlock.baseRangeMax]	[[mayor frecuencia]]
em_res_power	[spectralWindow.resolution]	[spectralWindow.resolution]
o_ucd	em.mm	em.mm
pol_states	[Source.stokesParameter[numStokes]]	[Source.stokesParameter[numStokes]]
facility_name	ALMA	ALMA
instrument_name	ALMA	ALMA

En la Tabla 3.6 se muestra el resultado de la investigación, la primera columna corresponde a las columnas de la clase *Observation*, la segunda columna indica de donde se obtienen los datos para llenar la los campos de la primera columna para el caso de los ASDM y la tercera para el caso de los FITS.

Los campos de ASDM o FITS que se encuentran encerrados entre corchetes indican de qué parte del ASDM son obtenidos, en caso de obtenerse a partir del FITS tienen un par de corchetes adicional.

Para poder llenar los campos de la clase *Observation* será necesario escribir un *script* que busque sobre las tablas del ASDM o sobre las cabeceras de los archivos FITS. Actualmente existen múltiples herramientas en el Paquete de Aplicaciones de Software Comunes de Astronomía (CASA, debido a sus siglas en inglés) [19].

3.3.2. Modelo de datos Characterisation de IVOA

Cuando se requiera entrar en más detalles al momento de realizar búsquedas, será necesario buscar en los metadatos almacenados por la clase Caracterización que consiste en la implementación del estándar de modelo de datos de IVOA *Characterisation*. Esta clase describe a través de distintos ejes el contenido de la observación, en el caso de ALMA existen 4 ejes relevantes: Espacial, Temporal, Espectral y Observable (flujo).

En la Figura 3.3, se encuentra el panorama general de los cuatro ejes relevantes del estándar de modelo de datos *Characterisation* para el caso de ALMA en ChiVO.

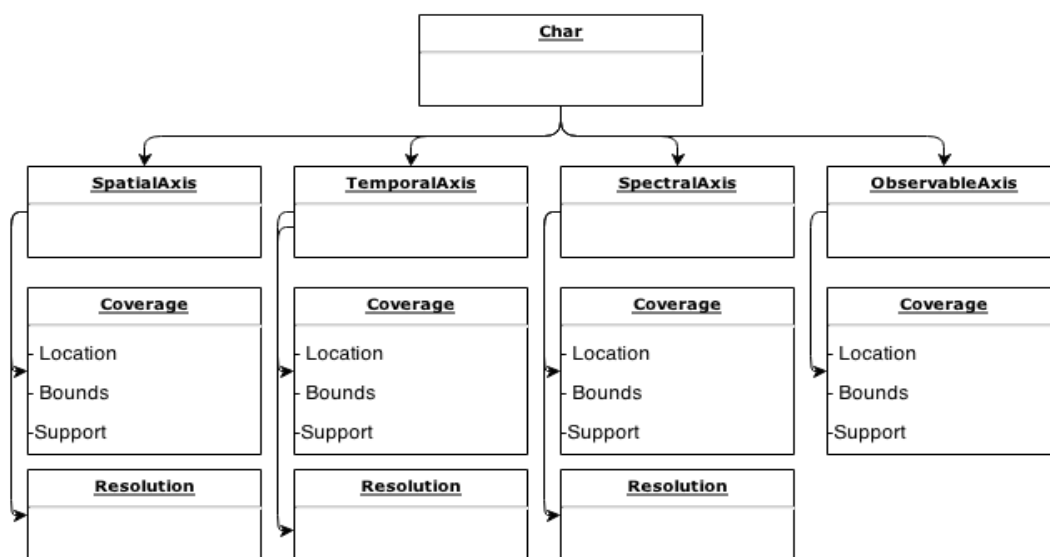


Figura 3.3. Modelo de Datos de alto nivel del estándar Characterisation para ALMA.

Las clases de la Figura 3.3 son explicadas a continuación:

AxisFrame (*SpatialAxis*, *TemporalAxis*, *SpectralAxis* y *ObservableAxis*): Esta clase contiene las propiedades principales del eje como el sistema de coordenadas utilizado, de si el eje es independiente del resto, unidades utilizadas, entre otros. Cabe destacar que el estado de calibración en el caso de los ejes es *CALIBRATED* si el eje se encuentra descrito por una unidad física, lo cual se cumple en este caso, para los cuatro ejes.

Coverage.Location: Esta subclase de *Coverage* o cobertura, describe el valor característico del eje. Por ejemplo, en el caso del eje espacial *Spatial.Coverage.Location* el valor característico sería la coordenada del punto central de la observación, o en el caso del eje espectral la frecuencia media.

Coverage.Bounds: Esta subclase de *Coverage* describe los valores mínimos y máximos del eje. Por ejemplo, en el caso del eje espacial *Spatial.Coverage.Bounds*

sería la coordenada mínima y máxima de la observación para el caso de coordenadas ecuatoriales corresponde a RA y DEC.

Coverage.Support: Conjunto de parámetros que describen de mejor manera los límites de la observación donde se encuentran los datos válidos de ésta. Por ejemplo, en el eje espacial *Spatial.Coverage.Support* describe el círculo de la observación con un punto central y un radio, ya que en el caso de ALMA la observación no es rectangular.

Resolution: Esta clase describe la resolución del eje.

Al igual que para el estándar *Observation Core* de IVOA, el estándar *Characterisation* también tiene requerimientos para que su implementación sea compatible con el ecosistema de los observatorios virtuales. Estos requerimientos son incluir al menos *Location* o *Bounds* en *Coverage* y con un valor numérico explícito en *Location*, *Bounds* y *Support* en caso de estar presentes. Además, todos los ejes presentes deben tener unidad y sistema de coordenadas, y utilizar las definiciones del STC para ello.

Cada eje tiene sus propias definiciones y acorde a las entrevistas realizadas al radioastrónomo Diego Mardones, se definieron en detalle cada eje y sus propiedades. A continuación se describe cada uno de ellos.

3.3.2.1. Eje Espacial

El eje espacial describe las coordenadas de la observación y el área observada, en la Figura 3.4 se muestran las clases involucradas.

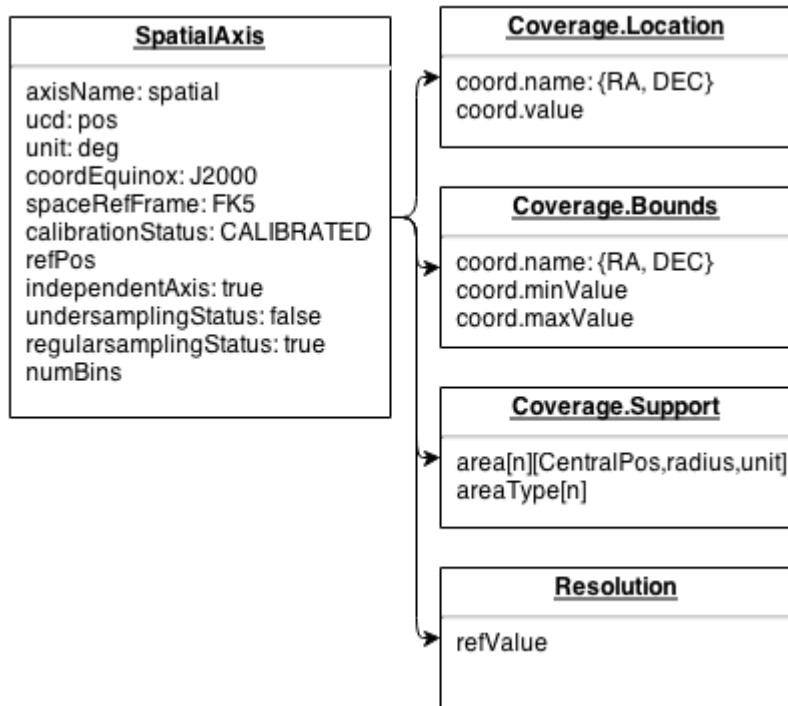


Figura 3.4. Eje espacial de la clase de caracterización.

SpatialAxis

Describe las propiedades principales del eje espacial. Se escoge el tipo de sistema de coordenadas que en este caso es el sistema ecuatorial con J2000, las coordenadas serán expresadas en grados debido a que es ampliamente utilizado en la comunidad astronómica, los datos contenidos en este eje se encuentran calibrados ya que tienen una unidad asociada, refPos hace referencia al punto medio, o sea, al centro de la observación. Además, este eje es independiente del resto, la cantidad de datos no es tan baja para la resolución del instrumento por lo que no se encuentra con *undersampling* y hay un espacio regular entre un pixel y otro, o sea es regular *sampling*.

Coverage.Location

En el caso de ALMA estos valores corresponden al punto central de la observación.

Coverage.Bounds

Corresponde al límite superior e inferior de las coordenadas de la observación, para el caso de ALMA son los valores mínimos y máximos en RA, DEC. Esta definición es genérica y definida por el estándar, formando un rectángulo alrededor de la coordenada central, sin embargo, en el caso de ALMA las observaciones son mejor representadas por un círculo, lo cual se describe en Support.

Coverage.Support

Debido a que las observaciones de ALMA son un cono proyectado sobre el cielo y no un rectángulo, es necesaria la clase Support, que describe todos los círculos de la observación contenidos en el ASDM. En el caso de ser un FITS, sólo habría un círculo.

Resolution

En esta clase se almacena la resolución espacial, para el caso del ASDM se utiliza $\frac{1.2 \cdot \lambda}{D}$, donde D corresponde a la separación máxima entre las antenas, si bien es una buena referencia, el valor de la resolución dependerá de qué distancia entre antenas se considera, pudiendo ser cualquier distancia entre la distancia mínima y la distancia máxima. En el caso de los FITS, se debería encontrar esta información en la cabecera.

3.3.2.2. Eje Temporal

El eje temporal describe todo lo relacionado con el tiempo de la observación, como la duración y fechas de observación, en la Figura 3.5 se muestran las clases involucradas en el eje.

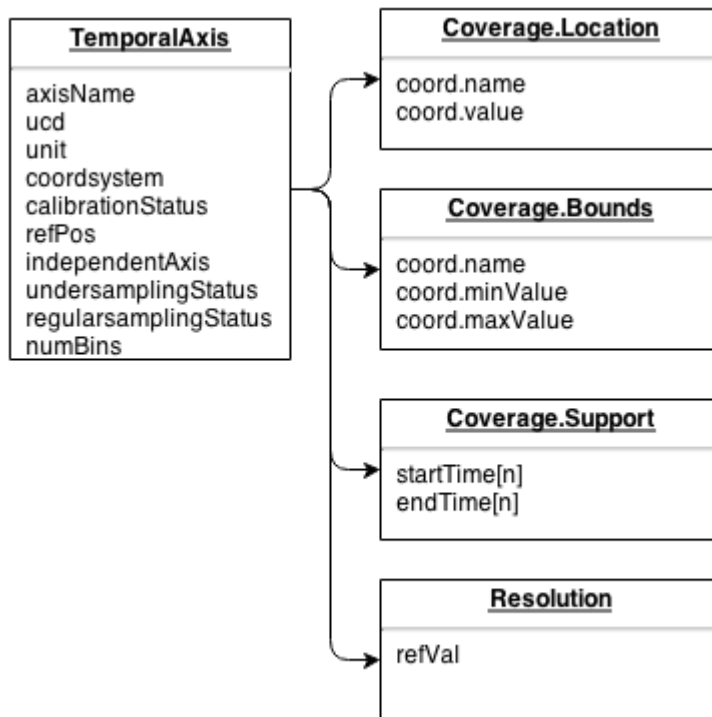


Figura 3.5. Eje temporal de la clase de caracterización.

TemporalAxis

Describe las propiedades principales del eje temporal. El sistema de coordenada corresponde al UTC, los datos se encuentran calibrados ya que son representados en segundos, la posición de referencia es topocéntrica y es un eje independiente. En el caso de ASDM se tiene numBins que representa la cantidad de observaciones, este número es luego utilizado en Support para describir el rango de tiempo de cada observación, para los FITS numBins no es necesario ya que un FITS corresponde a una observación con sólo un tiempo de inicio y uno de fin.

Coverage.Location

En el caso de los ASDM estos valores corresponden al tiempo medio entre el tiempo de inicio de la primera observación y el tiempo final de la última, en el caso de FITS sólo el tiempo medio entre el tiempo inicial y final.

Coverage.Bounds

La definición del estándar indica que *Bounds* tiene como valor mínimo el tiempo inicial de la primera observación y como valor máximo el valor del tiempo

final de la última observación. En el caso de FITS, es sólo el tiempo inicial y final de la observación.

Coverage.Support

Sólo aplica en el caso de ASDM ya que es necesario especificar todos los rangos de tiempo de las observaciones, esta cantidad está especificada en la clase *TemporalAxis* en *numBins*.

Resolution

Contiene la resolución temporal, en el caso de ALMA se podría utilizar el campo *interval* en la tabla *mainTable* del ASDM.

3.3.2.3. Eje Espectral

El eje espectral describe las frecuencias de la observación, en la Figura 3.6 se muestran las clases involucradas.

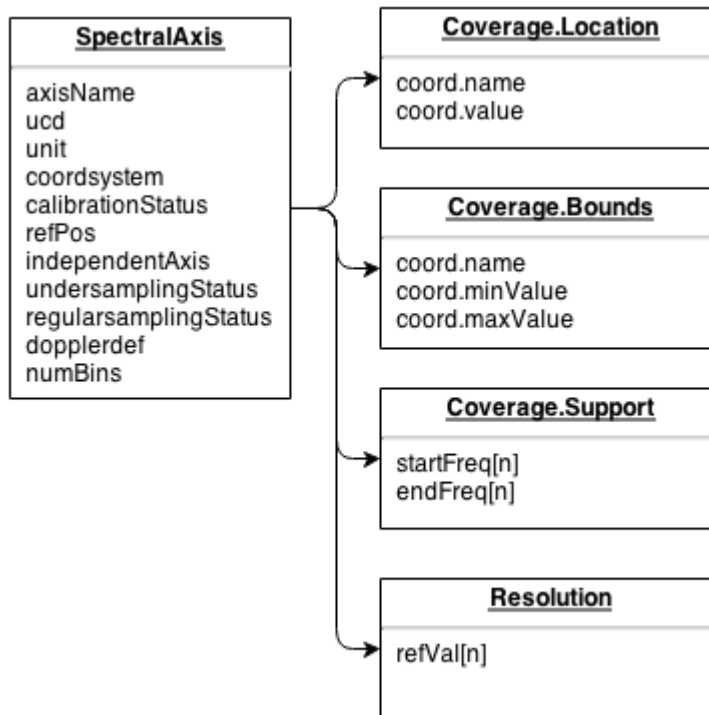


Figura 3.6. Eje espectral de la clase de caracterización.

SpectralAxis

Describe las propiedades principales del eje espectral. El sistema de coordenadas utilizado es el Sistema Internacional con la unidad utilizada para representar frecuencia, Hertz. Los datos en este eje se encuentran en estado de calibración *CALIBRATED* ya que son representados por la unidad Hz. La posición de referencia es la frecuencia media de todas las frecuencias observadas. Este eje es independiente del resto. La cantidad de datos no es tan baja para la resolución del instrumento por lo que no se encuentra con *undersampling* y hay un espacio regular entre los datos o sea, es *regular sampling*. Además, en este caso los valores de este eje se encuentran afectados por el efecto doppler, por lo que en *dopplerdef* se encuentra el valor de la velocidad en caso de representar datos extragalácticos cercanos, o z en caso de ser lejanos. En el caso de los ASDM, hay varios rangos de frecuencias contenidos, *numBins* indica la cantidad de rangos.

Coverage.Location

Para el caso de ASDM y FITS se utiliza la frecuencia central observada.

Coverage.Bounds

En el caso de los ASDM, los valores mínimos y máximos corresponden a al valor mínimo del rango menor y al valor máximo del rango mayor de frecuencias, respectivamente. En el caso de FITS, hay sólo un rango por lo que los valores mínimos y máximos corresponden al mínimo y máximo de dicho rango.

Coverage.Support

Esta clase sólo es necesaria para el caso de ASDM ya que indica con detalle todos los rangos de frecuencias contenidos en el ASDM.

Resolution

En el caso de ALMA la resolución espectral se encuentra en la descripción de cada *Spectral Window* (ventana espectral) y corresponde al ancho de cada canal expresado en frecuencia. En el ASDM se encuentra en la tabla

SpectralWindow.Resolution y debe indicarse una resolución para cada rango de frecuencia contenido en el ASDM. Para los FITS sólo uno ya que sólo hay un rango de frecuencias y corresponden a un mismo canal.

3.3.2.4. Eje Observable

El eje observable describe el flujo de la observación, en la Figura 3.7 se muestran las clases involucradas.

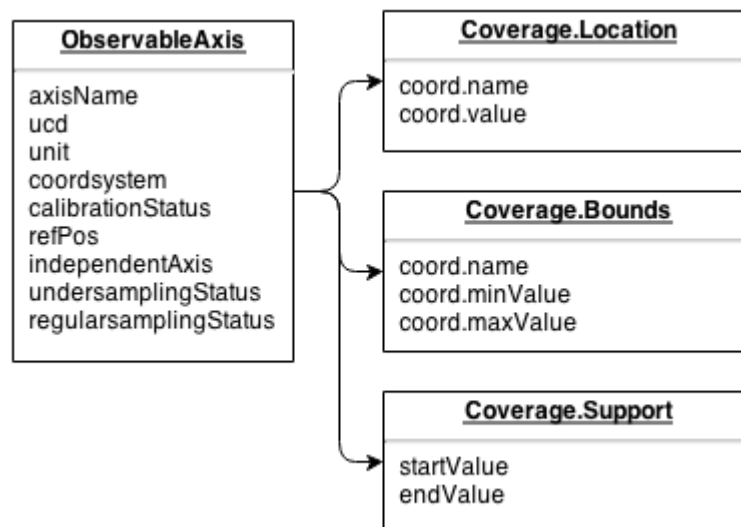


Figura 3.7. Eje observable de la clase de caracterización.

ObservableAxis

Describe las propiedades principales del eje observable o del flujo. Las unidades utilizadas pueden ser Jy, Jy/beam o K según el sistema definido por la *International Astronomical Union*, ya que este eje tiene una unidad definida tiene un estado de calibración *CALIBRATED* este eje independiente del resto.

Coverage.Location

El valor contenido en esta clase es el flujo medio.

Coverage.Bounds

En esta clase se indican los valores mínimos y máximos en el flujo. Para el caso del flujo mínimo se utiliza el error en el flujo, o sea, el sigma o RMS. Para el caso del valor máximo, en el caso de los ASDM se debe recorrer el arreglo *flux* ubicado en la tabla *source*, hasta encontrar el valor máximo.

Coverage.Support

Esta clase contiene información más detallada respecto al eje observable o de flujo, pero no aporta mayor información ya que es una definición, el valor mínimo es tres veces el error en el flujo, o sea, tres veces el valor mínimo de la clase *Bounds*, el valor máximo del *Support* es el mismo que expresado en la clase *Bounds*.

3.4 Arquitectura

En los capítulos anteriores se han descrito las necesidades de los radioastrónomos chilenos, los requerimientos y casos de uso obtenidos a partir de ello y los modelos de datos compatibles con los estándares de IVOA. Ahora para poder entender mejor cómo será ChiVO y terminar el diseño de este Observatorio Virtual, es necesario detallar la arquitectura de este sistema. En la sección 2.1 del estado del arte se explica la arquitectura propuesta por IVOA, y en la Figura 2.1 se puede ver en detalle un diagrama que la representa. En la Figura 3.8, se encuentra un diagrama de la arquitectura que tendrá ChiVO.

3.4.1. Capa de clientes

Esta capa representa al usuario final y cómo facilita la comunicación entre el usuario y los datos. En esta capa el usuario realiza consultas a través de los protocolos de acceso ofrecidos por ChiVO o a través de un formulario avanzado, utilizando aplicaciones compatibles con el ecosistema OV y el portal web, respectivamente. Una vez realizada la consulta, el sistema le retornará al usuario una lista que describe objetos u observaciones encontrados (metadatos) y podrá acceder a ellos a través de un enlace de descarga asociado a cada resultado.

Cabe destacar que gracias a la separación por capas, se logra flexibilidad y escalabilidad en el sistema, para que independiente de una capa y otra, puedan interactuar nuevas aplicaciones y portales con ChiVO, así como la adición de nuevas fuentes de información a parte de ALMA.

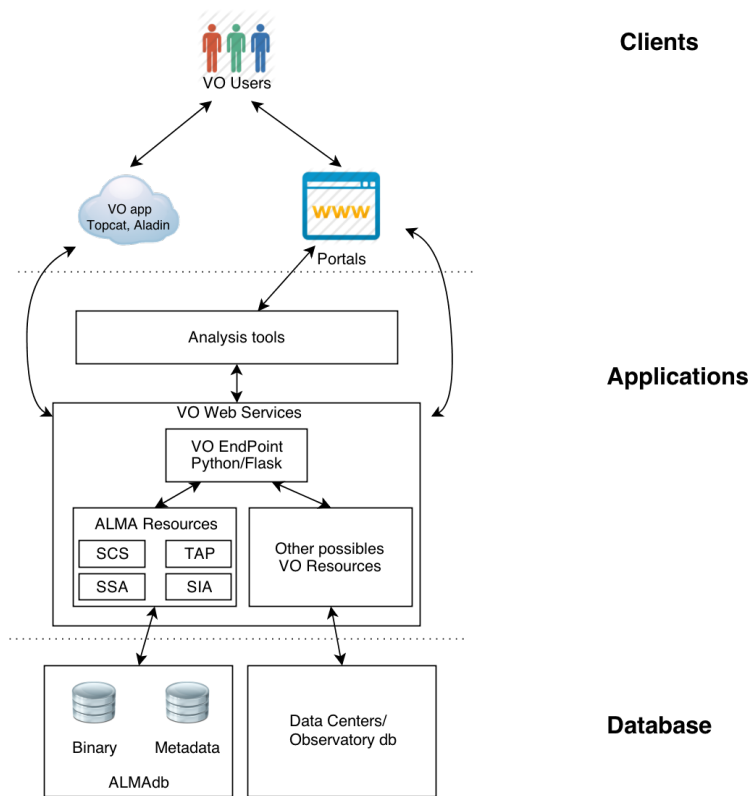


Figura 3.8. Arquitectura de ChiVO.

Las consultas son recibidas por ChiVO a través de su *endpoint* de datos que recibe consultas en HTTP, GET o POST, ante lo cual el *endpoint* retorna la lista de resultados en una tabla en el formato VOTable. Para el caso del portal web, el VOTable será mostrado al usuario a través de una herramienta web que permite la manipulación simple y eficiente de VOTables llamada VOVView.

3.4.2. Capa de Aplicaciones

En esta capa se encuentran los programas que procesan las consultas entre los usuarios y los datos. Para ello es necesario un Servicio Web compatible con el ecosistema de OV que sea compatible al menos con el TAP para acceder al modelo de datos de ChiVO, el protocolo para realizar búsquedas cónicas *Simple Cone Search* (SCS) de IVOA, el protocolo para realizar acceder a datos espectrales *Simple*

Spectral Access (SSA) de IVOA y el protocolo de acceso a imágenes *Simple Image Access* (SIA) de IVOA. Cada estándar de IVOA requiere un mínimo de su propia implementación para ser compatible con el ecosistema de OV, en el caso de estos protocolos de acceso sólo es necesaria la recepción de consultas HTTP básicas junto a los parámetros de búsqueda requeridos.

La caja que representa a las herramientas de análisis es fundamental en la eficiencia de ChiVO, esto es debido a que los datos a analizar por los astrónomos suelen tener un gran tamaño y es costosa su transferencia, este problema se resuelve acercando las herramientas de análisis y procesamiento al lugar donde están almacenados los datos a procesar. Actualmente en Chile se utiliza este sistema en el Centro de Modelado Matemático de la Universidad de Chile, donde al usuario se le entrega un espacio donde puede almacenar datos a procesar y puede ejecutar *scripts* sobre éstos.

Debido a que más adelante será necesario ofrecer búsquedas por otros datos que no sólo provengan de ALMA, es necesaria cierta abstracción al momento de implementar esta capa, ya que debe permitir a futuro interactuar con nuevos recursos pero siempre manteniendo la compatibilidad de IVOA para que sea utilizable por aplicaciones compatibles con el ecosistema de IVOA.

3.4.3. Capa de Datos

En esta capa se encuentran los recursos que tienen los datos y metadatos. Para ello aquí es necesaria la implementación de los modelos de datos presentados en 3.3.1 y 3.3.2, los estándares de modelos de datos *Observation Core* y *Characterisation*.

Capítulo 4: Conclusiones

4.1. Análisis de los resultados obtenidos

La lista de requerimientos es el primer resultado obtenido, éstos se construyeron en base a las necesidades de los astrónomos a través de entrevistas con varios astrónomos a nivel nacional.

A partir de los requerimientos, se obtuvieron casos de uso para que así ChiVO pueda ser desarrollado. Gracias a que fue validado por varios radioastrónomos nacionales, los casos de uso satisfacen las necesidades nacionales por parte de la comunidad científica astronómica. De estos, se puede apreciar que no es necesario construir herramientas nuevas, la mayoría de los casos de uso ya existen a nivel internacional en otros Observatorios Virtuales, pero sin el acceso a los datos astronómicos de los chilenos. Es por ello que este proyecto es factible a corto o mediano plazo, la búsqueda sobre datos chilenos pronto será una realidad.

El resultado más importante es el obtenido con los modelos de datos para realizar búsquedas, lo cual es la base de ChiVO. Para ello, se proponen dos implementaciones de estándares de IVOA: Modelos de datos *Observation Core* y *Characterisation*, que juntos permiten a los astrónomos descubrir y analizar a través de metadatos. El mayor logro fue la concretización de lo requerido por IVOA, ALMA y la comunidad chilena en un modelo de datos que está listo para implementar, sin sacrificar flexibilidad ni escalabilidad. Además, gracias a reuniones con miembros de las 3 partes, se validó el modelo de datos realizando una lista de donde obtener los datos para poblar la futura base de datos de ChiVO fácilmente.

Finalmente se presenta una arquitectura para ChiVO en base a lo sugerido por IVOA, es básica pero muy flexible, lo cual es muy importante para ChiVO ya ALMA es sólo el comienzo.

4.2. Futuros trabajos

Esta memoria está aún lejos de transformarse en ChiVO, pero es el primer paso en conjunto con la reciente integración de ChiVO a IVOA. Los siguientes pasos son un gran desafío para Chile pero gracias al equipo que se ha formado de varias Universidades chilenas, es posible.

Lo que queda a futuro son parte de un proyecto de creación de *software*, donde esta memoria cumple la parte de toma de requerimientos y diseño. A continuación se deben implementar los resultados de esta memoria y seguir adelante con las necesidades que la comunidad científica astronómica requiera.

Es crítico también mantenerse constantemente participando en IVOA, ya que ahora Chile es miembro, tiene el gran compromiso de aportar con la construcción de este gran Observatorio Virtual.

Finalmente, y no menos importante por ello, ChiVO no será nada si no es respaldado por toda la comunidad. Es fundamental que se realice un fuerte trabajo con los estudiantes de astronomía mostrándoles las ventajas que obtienen con un Observatorio Virtual. Aprovechar también de realizar algo muy importante, llegar a los colegios. Es casi inexistente esta ciencia en el colegio, y debería motivarse a través de charlas o planetarios (que en Chile tienen mucho por crecer).

Bibliografía

- [1] Real Academia Española. Definición de astronomía [en línea]
<<http://lema.rae.es/drae/?val=astronomía>> [consulta: 26 noviembre 2013].
- [2] Real Academia Española. Definición de cosmología [en línea]
<<http://lema.rae.es/drae/?val=cosmología>> [consulta: 26 noviembre 2013].
- [3] BERRIMAN, Bruce G. y GROOM, Steven L. *How Will astronomy archives Survive the data tsunami? En revista Communications of the ACM*, 54(12):52-58, Diciembre 2011.
- [4] *The Virtual Observatory – A New Era for Astronomy* por Carvalho R., et al. *Journal of Computational Interdisciplinary Sciences*, 1(3):187-206, 2010.
- [5] IVOA. ¿Qué es IVOA? [en línea]
<<http://www.ivoa.net/about/what-is-ivoa.html>> [consulta: 26 noviembre 2013].
- [6] VAO. ¿Qué es un Observatorio Virtual? [en línea]
<<http://www.virtualobservatory.org/whatis>> [consulta: 26 noviembre 2013].
- [7] WHITE, Tom. *Hadoop: The Definitive Guide*. 3era edición, 2012.
- [8] IVOA. Arquitectura de un Observatorio Virtual [en línea]
<<http://www.ivoa.net/Documents/Notes/IVOAArch/IVOAArch-20040615.html>>
[consulta: 26 noviembre 2013].
- [9] WILLIAMS, Roy, et al. Pauta de participación [en línea]
<<http://www.ivoa.net/documents/Notes/IVOA/IVOAParticipation-1.1-20100707.pdf>> [consulta: 26 noviembre 2013].
- [10] MARINA de E.E.U.U. *The Astronomical Almanac Online*: Observatorios por país [en línea] <<http://asa.usno.navy.mil/>> [consulta: 26 noviembre 2013].
- [11] BARBA, Rodolfo, et al. La necesidad en Chile de una infraestructura tecnológica colaborativa de apoyo a la investigación astronómica. Programa Bicentenario de Ciencia y Tecnología, Noviembre 2008.
- [12] *Astronomy, technology, industry. Roadmap for the fostering of Technology Development and Innovation in the Field of Astronomy in Chile*. Conicyt, Santiago, Chile, 30 Octubre 2012.

- [13] CDS. SIMBAD: Base de datos astronómica [en línea]
<<http://simbad.u-strasbg.fr>> [consulta: 26 noviembre 2013].
- [14] CDS, Resolvedor de nombres astronómicos [en línea]
<<http://cds.u-strasbg.fr/resources/doku.php?id=namerresolver>> [consulta: 26 noviembre 2013].
- [15] CALTECH. Base de datos astronómica de extragaláctica [en línea]
<<http://ned.ipac.caltech.edu/>> [consulta: 26 noviembre 2013].
- [16] Mazzarella, Joseph, Equipo NED. *Chapter 18: Web-based Tools—NED VO Service*. En *The National Virtual Observatory Book*, 2008, 165-178.
- [17] CDS. Servicio VizieR de búsqueda de catálogos [en línea]
<<http://vizier.u-strasbg.fr/viz-bin/VizieR>> [consulta: 26 noviembre 2013].
- [18] IVOA. Arquitectura de IVOA [en línea]
<<http://www.ivoa.net/documents/Notes/IVOAArchitecture/20101123/IVOAArchitecture-1.0-20101123.pdf>> [consulta: 26 noviembre 2013].
- [19] NRAO, *Common Astronomy Software Applications* [en línea]
<<http://casa.nrao.edu>> [consulta: 26 noviembre 2013].
- [20] CADC. Resolvedor de nombres [en línea]
<<http://www3.cadc-ccda.hia-ih.nrc-cnrc.gc.ca/NameResolver>> [consulta: 26 noviembre 2013].
- [21] NRAO. Resolvedor de moléculas y frecuencias [en línea]
<<http://www.cv.nrao.edu/php/splat/>> [consulta: 26 noviembre 2013].
- [22] VAO. Nota sobre interfaz SLAP [en línea]
<<http://www.cv.nrao.edu/php/splat/SLAPNotes2.html>> [consulta: 26 noviembre 2013].
- [23] SAO, NASA. Sistema de Datos astrofisico [en línea]
<<http://adsabs.harvard.edu>> [consulta: 26 noviembre 2013].
- [24] ADS. Lista de palabras claves de parámetros para consultas a ADS [en línea]
<http://ads.harvard.edu/cgi-bin/get_field_names.pl> [consulta: 26 noviembre 2013].
- [25] ADS. Acceso a servicios de ADS con formularios embebidos [en línea]

<http://doc.adsabs.harvard.edu/abs_doc/help_pages/linking.html#Embedded_Queries> [consulta: 26 noviembre 2013].

[26] ADS. Código fuente de la API para acceder a los servicios de ADS [en línea]
<<https://github.com/adsabs/adsabs-dev-api>> [consulta: 26 noviembre 2013].

[27] NASA. Definición de FITS [en línea]
<<http://fits.gsfc.nasa.gov>> [consulta: 26 noviembre 2013].

[28] Pence, W.D. *Definition of the Flexible Transport System (FITS)*, versión 3.0. Astronomy & Astrophysics. 524(A42), Noviembre 2010.

[29] WEST, Matthew. *Developing High Quality Data Models*. 1era edición, 2011.

[30] SANTANDER, Juan de Dios. Bases de Datos Multimedia para Radioastronomía: RADAMS y DSS-63. Memoria (Diploma de Estudios Avanzados en el programa de doctorado Tecnologías Multimedia). Andalucía, España. Instituto de Astrofísica de Andalucía, 2006. 98 h.

[31] IVOA. Estándar del Table Access Protocol [en línea]
<<http://www.ivoa.net/documents/TAP>> [consulta: 26 noviembre 2013].

[32] IVOA. Estándar del Simple Spectral Access [en línea]
<<http://www.ivoa.net/documents/SSA>> [consulta: 26 noviembre 2013].

[33] IVOA. Estándar del Observation Core [en línea]
<www.ivoa.net/documents/ObsCore> [consulta: 26 noviembre 2013].

[34] STUART, Lowe. LookUP resolvidor de nombres [en línea]
<<http://www.strudel.org.uk/lookUP>> [consulta: 26 noviembre 2013].

Anexo A: Estándares de IVOA

En este anexo se describen los principales estándares de IVOA involucrados con el modelo de datos y el acceso a éstos. Además, se mencionan de forma general otros estándares de IVOA que serán relevantes para ChiVO más adelante. Estos estándares son obtenidos del documento oficial de IVOA [18] que describe la arquitectura de éstos.

A.1. Estándares de Protocolos de obtención y acceso a datos

La mayoría de los protocolos de acceso a datos de IVOA usan VOTable para intercambio de metadatos y STC, UCD, Utypes y Units para descripción de los metadatos contenidos en los VOTables.

A.1.1. SIAP (*Simple Image Access Protocol*)

Define una interfaz para consultar y recibir datos de imágenes de una variedad de repositorios de imágenes. El servicio retorna una serie de imágenes candidatas formateadas como un VOTable. Cada candidato tiene una referencia URL de acceso en donde está disponible la imagen.

A.1.2. SCS (*Simple Cone Search*)

Define un protocolo simple de consultas para recibir registros desde un catálogo de recursos astronómicos. El servicio recibe una consulta que describe una posición en el cielo y una distancia angular definiendo así un cono en el cielo, el servicio retorna lo que encuentre en dicho cono.

A.1.3. SSAP (*Simple Spectra Access Protocol*)

Define una interfaz uniforme para descubrimiento remoto y acceso a un espectro en una dimensión que provienen de colecciones de datos y metadatos. SSAP está basado en el *Spectrum Data Model*, el cual es capaz de describir algunos datos como series de tiempo y distribuciones de energía espectrales (SED), así como también espectros de una dimensión. SSAP puede ser usado por aplicaciones de OV para acceder a los recursos espectrales asociados.

A.1.4. SLAP (*Simple Line Access Protocol*)

Define una interfaz uniforme para el descubrimiento y acceso de transiciones de líneas espectrales desde colecciones de datos y metadatos. SLAP está basado en el SSLDM (*Simple Spectral Line Data Model*).

A.1.5. TAP (*Table Access Protocol*)

Define una interfaz para consultar datos de la tabla general. El acceso es provisto tanto para base de datos y la tabla de metadatos, como también para los datos propios de la tabla. Se incluye soporte a múltiples lenguajes de consulta tales como ADQL (*Astronomical Data Query Language*) y PQL (*Parametrized Query Language*). También soporta consultas sincrónicas y asincrónicas. Soporta tres tipos de consultas: consultas de datos, consultas de metadatos y consultas VOSI.

A.1.6. SEAP (*Simple Event Access Protocol*)

Define una interfaz uniforme para descubrimiento remoto y acceso de eventos astronómicos provenientes de colecciones de datos y metadatos. Está basado en el VOEvent *Data Model* que define el contenido y significado de un paquete de información estándar para representar, transmitir, publicar y lograr el descubrimiento de un evento celestial fugaz. En 2010 continúa en desarrollo.

A.1.7. SimDAL (*Simulations Data Access Layer*)

Define una forma uniforme para acceso a simulaciones y productos teóricos. Trabaja con el *Simulation Data Model* (SimDM).

A.1.8. FAP (*Footprint Access Protocol*)

Define una interfaz uniforme para el descubrimiento remoto y acceso de *footprints* provenientes de colecciones de datos y metadatos. Un *footprint* es una descripción poligonal de la extensión espacial de un conjunto de datos. En 2010 se encuentra en desarrollo.

A.2. Núcleo del OV: Estándares de Modelos de Datos

A.2.1. STC (*Space Time Coordinate metadata*)

Este estándar provee un *framework* para describir metadatos espaciales y temporales. Utilizado en la mayoría de los estándares IVOA que hacen referencia a metadatos espaciales y/o temporales. STC puede ser serializado con un VOTable.

A.2.2. Units

Son convenciones de nombres y representaciones. El VO Units DM tiene como finalidad definir prácticas comunes en la manipulación de unidades en metadatos astronómicos y definirá significados de representación consistente dentro de los servicios VO. En 2010 el estándar *Units* DM está en desarrollo.

A.2.3. Utypes

Son nombres que definen inequívocamente un elemento de un modelo de datos que representa una pieza de datos en el VO. *Utypes* DM serán usados por todos los modelos de datos IVOA, así también en el contexto de los Protocolos de Acceso de Datos y Lenguajes de consulta de los OV. Está relacionado con los *Units* DM y será usado para el intercambio de metadatos a través del VOTable.

A.2.4. SpectrumDM

Presenta un modelo de datos que describe la estructura de conjuntos de datos espectrofotométricos con coordenadas espectrales y temporales y metadatos asociados. Puede ser utilizado para representar espectros, series de tiempos, segmentos de distribución de energía espectral y otras asociaciones espectrales y espaciales. Usado con SSAP.

A.2.5. SSLDM (*Simple Spectral Line DM*)

Presenta un modelo de datos que describe transiciones de líneas espectrales. El SSLDM es usado con el SLAP.

A.2.6. CharDM (*Characterization Data Model*)

Define y organiza todos los metadatos necesarios para describir como un conjunto de datos ocupa un espacio físico multidimensional cuantitativamente y, cuando es relevante, cualitativamente. El modelo de datos se centra en los ejes usados para

trazar este espacio, incluyendo pero no limitándose a los ejes espaciales (2D), temporales, espectrales, así como el eje Observable, o algún otro eje físico. CharDM es una abstracción que puede ser usada para derivar en una descripción estructurada de datos y así facilitar el descubrimiento y la interpretación científica.

A.2.7. ObsCoreDM (*Observation Core Data Model*)

Define los componentes centrales de todos los metadatos de consultas que tienen un rol en el descubrimiento de observaciones. Se centra en un conjunto razonable de descriptores de metadatos disponibles y así permite un costo razonable de implementación para los proveedores de datos. A menudo ha sido referido como el proyecto ObsTAP.

A.2.8. ObsProvDM (*Observation and Provenance DM*)

Define todos los componentes de una observación, incluyendo información acerca de su procedencia, como ha sido procesado, sobre qué condición de observación ha sido llevado a cabo, etc. Algunas definiciones detalladas y las potenciales independencias con otros estándares IVOA necesitan ser definidas. Es una extensión del modelo ObsCoreDM y probablemente hará referencia al CharDM.

A.2.9. PhotDM (*Photometry DM*)

Describe filtros fotométricos a través de un modelo de datos simple con la finalidad de crear protocolos para acceder a datos fotométricos en magnitudes. PhotDM podría ser usado en conjunto con otros estándares IVOA, tales como SSAP o TAP. Phot hace referencia a CharDM, SpectrumDM y a la parte de procedencia de ObsProvDM.

A.2.10. SimDM (*Simulations Data Model*)

Define y organiza todos los metadatos necesarios para describir conjuntos de datos de simulaciones. Puede ser usado en conjunto con SimDAL. En 2010 este estándar está en desarrollo.

A.2.11. VOEvent

Define el contenido y significado de un paquete de información estándar para representar, transmitir, publicar y lograr el descubrimiento de un evento celestial fugaz. Lo que implica que se está solicitando un seguimiento oportuno.

VOEvent está enfocado en el reporte de eventos de fotones, pero también pueden ser reportados eventos mediados por fenómenos dispares, tales como neutrinos, ondas gravitacionales, y ráfagas de partículas solares o atmosféricas.

A.3. Núcleo del VO: Estándares de Lenguajes de Consulta

A.3.1. ADQL (*Astronomical Data Query Language*)

Basado en SQL (*Structured Query Language*), es un lenguaje que soporta consultas específicas para la astronomía. Las funciones geométricas de ADQL están definidas por el estándar STC.

A.3.2. PQL (*Parametrized Query Language*)

Es un lenguaje para representar consultas astronómicas simples publicadas a servicios VO.

Anexo B: Descripción detallada de columnas del *Observation Core*.

B.1. *Data Product Type* `_(dataprodukt_type)_`

La columna *dataprodukt_type* contiene una cadena de caracteres simple que describe la naturaleza primaria del producto de datos. Debe asumir uno de los siguientes valores: **image**, **cube**, **spectrum**, **sed**, **timeseries**, **visibility** o **event**.

- **image**: Imagen astronómica 2D con dos ejes espaciales (FITS).
- **cube**: Imagen astronómica multidimensional con 3 o más ejes. (FITS multidimensional).
- **spectrum**: Conjunto de datos con la cobertura espectral como atributo primario. (espectro 1D).
- **sed**: distribución de energía espectral (datos de múltiples observaciones).
- **timeseries**: es un arreglo de una dimensión que presenta alguna cantidad como función del tiempo (una curva de luz).
- **visibility**: Conjunto de visibilidades (en radio) de algún tipo (datos de ALMA).
- **event**: Conjunto recuento de eventos de algún tipo.

El valor NULL es permitido, pero solo si ninguno de los valores propuestos puede ser usado para describir al conjunto de datos. Puede usarse el campo opcional *dataprodukt_subtype* para definir con mayor precisión la naturaleza del conjunto de datos y para el formato por los atributos *obs_title* y *access_format*. Los valores deben estar escritos en minúsculas. Especificar este campo junto con la cobertura espacial y espectral deseada será suficiente para descubrir datos de interés en la mayor parte de los casos.

Uso: `select * from ivoa.ObsCore where dataprodukt_type='image'` retorna solo datos image.

B.2. Calibration Level *_(calib_level)_*

La columna *calib_level* muestra el nivel de procesamiento de calibración que ha sido aplicado para crear el producto de datos. Los valores en esta columna no deben ser nulos. Queda en los proveedores de datos decidir los casos ambiguos. Puede tener los valores enteros 0, 1, 2 y 3:

Nivel 0: datos instrumentales en bruto, en formato definido por un propietario o un proveedor de datos interno, que necesita herramientas específicas de los instrumentos para ser manejados.

Nivel 1: datos instrumentales en formato estándar (FITS, VOTable, SDFITS, ASDM, etc.) los cuales pueden ser manipulados por paquetes astronómicos estándar.

Nivel 2: datos científicos calibrados, listos con la firma de instrumento removida.

Nivel 3: productos de datos mejorados tales como mosaicos, campos de estudio elaborados. Puede representar la combinación de datos de múltiples observaciones.

Uso: `select * from ivoa.ObsCore where calib_level > 2` retorna productos de datos mejorados.

B.3. Collection Name *_(obs_collection)_*

La columna *obs_collection* identifica la colección de datos para los cuales pertenece el producto de datos. Una colección de datos (*data collection*) es cualquier colección de un conjunto de datos que es semejante en alguna manera. Colecciones de datos típicas pueden ser datos de un telescopio, instrumento, o una inspección (*survey*). El valor es el nombre corto registrado de la colección de datos o el identificador IVOA completo para la colección o el nombre corto del proveedor de datos para la colección. A menudo el nombre de la colección será definido por el nombre del

instrumento que genera los datos. In ese caso se sugiere que el nombre de la colección sea un *string* compuesto por el *facility name*, seguido por un *slash* y luego el nombre del instrumento.

Ejemplo: HST/WFPC2, VLT/FORS2, CHANDRA/ACIS-S

En algunos casos no es necesario utilizar el nombre del instrumento porque los *data products* usan datos desde diferentes instrumentos, instituciones o por otras razones.

En la práctica no es un campo precisamente definido. Lo importante para el proveedor de datos del nombre de la colección es que sea familiar para los astrónomos y sea discriminativo para apuntar a conjunto de datos de interés.

Los valores no deben ser NULL.

B.4. *Observation Identifier* _(obs_id)_

La columna *obs_id* define un identificador único para la observación. En el caso de que múltiples productos de datos estén disponibles para una observación, cada producto que lo compone llevará el mismo valor *obs_id*. El valor de *obs_id* retornado para una observación archivada debe permanecer idéntico a través del tiempo para futuras referencias.

En el caso de algunos *data products* avanzados (con *calib_level*=3), el *data product* puede ser el resultado de la combinación de múltiples observaciones primarias, entonces un nuevo identificador de observación único debe ser asignado. Si varios *data products* se desprenden de estas observaciones todos deben llevar el mismo valor de *obs_id*.

Los valores en esta columna no deben ser NULL.

B.5. *Publisher Dataset Identifier* _(obs_publisher_did)_

La columna *obs_publisher_did* contiene el identificador del conjunto de datos IVOA para el producto de datos publicado. **Este valor debe ser único dentro del espacio**

de nombres controlado por el publicador (*publisher*) del conjunto de datos (*data center*). El valor es único si cada publisher tiene un *publisher ID* único registrado en IVOA. También puede tener más de un valor si este es publicado en más de una ubicación.

El valor retornado de *obs_publisher_did* para un data product estático se debería mantener inalterable con el paso del tiempo.

Los valores de la columna *obs_publisher_did* no deben ser nulos.

B.6. *Access URL* _(Access_url)_

La columna *access_url* contiene la URL que puede ser usada para descargar el data product (como archivo de algún tipo).

Se especifica el tipo de datos como CLOB (*character large object*) en el servicio TAP para que los usuarios sepan que solo puede usar la columna *access_url* en una sentencia SELECT de una consulta. Los usuarios no pueden especificar esta columna como parte de una sentencia WHERE.

El acceso URL no está garantizado para permanecer válido e inalterable de manera indefinida. Después de unas semanas se debe ejecutar una consulta (por ejemplo usando *obs_publisher_did*) para obtener un acceso URL reciente.

B.7. *Access Format* _(access_format)_

La columna *access_format* especifica el formato del data product si es descargado como un archivo. Este campo del modelo de datos es importante tanto para el descubrimiento de datos, como para el cliente que evaluará si será capaz de usar los datos si son descargados.

El valor para esta columna puede ser un tipo MIME, un tipo MIME estándar, un tipo MIME extendido o un nuevo tipo MIME personalizado definido por el data provider. Los tipos personalizados son del tipo “application/x-<tipo>”. El prefijo x es

requerido para cualquier cosa que no sea un tipo MIME estándar registrado. Puede ser usado e cualquier formato de archivo, incluyendo formatos de archivos binarios personalizados.

B.8. *Estimated Download Size* _(access_estsize)_

La columna *access_estsize* contiene el tamaño aproximado (en kilobytes) del archivo disponible vía *access_url*. Es importante para saber si los conjuntos de datos son muy grandes.

B.9. *Target Name* _(target_name)_

La columna *target_name* contiene el nombre del objetivo de la observación, si lo hay. Es típicamente el nombre del objeto astronómico, pero podría ser el nombre de un campo de estudio.

El nombre del objetivo es más útil como dato de salida, para identificar el objetivo de una observación para el usuario. En las consultas es mejor referirse a los objetos astronómicos por posición, convirtiendo el nombre en coordenadas si es posible.

B.10. *Cental Coordinates* _(s_ra,s_dec)_

El sistema de coordenadas en los que están expresadas las coordenadas es ICRS. La columna *s_ra* especifica el ICSR *Right Ascencion* del centro de la observación. La columna *s_dec* especifica el ICSR *Declination* del centro de la observación.

B.11. *Spatial Exent* _(s_fov)_

La columna *s_fov* contiene el tamaño aproximado de la región cubierta por el data product. Para regiones circulares corresponde al diámetro. Para otros *data products* corresponde a un valor lo suficientemente grande como para contener toda el área de observación. En observaciones donde no están bien definidos los límites (por ejemplo en observaciones de alta energía) un valor característico debe ser ingresado.

El atributo *s_fov* provee una forma simple de caracterizar y usar la cobertura espacial aproximada del *data product*. La cobertura puede ser definida más precisamente usando el atributo *s_region*.

B.12. *Spatial Coverage* *_(s_region)_*

La columna *s_region* puede ser usada para especificar precisamente la cobertura espacial de un *data product*.

Los servicios deberían especificar *s_region* mientras sea posible para especificar más precisamente la cobertura espacial de una observación.

B.13. *Spatial Resolution* *_(s_resolution)_*

La columna *s_resolution* especifica un valor de referencia escogido por el proveedor de datos para la resolución espacial estimada del *data product* en arco segundos. Esto se refiere al elemento espacial más pequeño en la señal observada. En el caso en que la resolución espacial varía a través del campo, la mejor resolución espacial debería ser especificada. En casos donde el muestreo de la frecuencia espacial es complejo (por ejemplo interferometría) un valor típico de resolución espacial estimado debería ser especificado.

B.14. *Time Bounds* *_(t_min, t_max)_*

La columnas *t_min* y *t_max* contienen los tiempos inicial y final de la observación especificados en MJD. En el caso de combinación de múltiples estructuras, el tiempo mínimo será el menor de todos los tiempos *t_min* y el tiempo máximo el mayor de todos los tiempos *t_max*.

B.15. *Exposure Time* *_(t_exptime)_*

La columna *t_exptime* contiene el tiempo de exposición. Para exposiciones simples es $t_{max} - t_{min}$. Para datos donde el detector no está activo en todo momento $t_{exptime} < t_{max} - t_{min}$. Para datos donde no es constante en el tiempo, suele usarse el tiempo medio de exposición. En algunos casos el tiempo de exposición es

un indicador de la sensibilidad relativa (profundidad) dentro de una colección de datos.

B.16. *Time Resolution* _(t_resolution)_

La columna *t_resolution* es el intervalo mínimo interpretable entre dos puntos a lo largo del eje de tiempo. Puede ser un promedio o un valor representativo. Para productos que no tienen muestreo a lo largo del eje de tiempo, puede ser especificado el tiempo de exposición o un valor nulo.

B.17. *Spectral Bounds* _(em_min, em_max)_

Las columnas *em_min* y *em_max* especifican los valores espectrales mínimo y máximo observado, expresados como una longitud de onda de vacío en metros.

B.18. *Spectral Resolving Power* _(em_res_power)_

La columna *em_res_power* contiene el poder de resolución espectral típico o característico definido como $(\lambda / \Delta \lambda)$. Es un valor adimensional.

B.19. *Observable Axis Description* _(o_ucd)_

La columna *o_ucd* especifica un UCD (*Undefined Content Descriptor*) describiendo la naturaleza de lo observable dentro del data product. Cualquier valor UCD válido es permitido. Si no hay un valor apropiado el campo debería ser dejado NULL.

B.20. Columnas adicionales

Los proveedores de datos pueden incluir columnas adicionales en la tabla *ivoa.ObsCore* para exponer metadatos adicionales. Estas columnas deben ser descritas en el *TAP_SCHEMA.columns* y en la salida desde los recursos VOSI-tables.

Anexo C: Observatorios Virtuales alrededor del mundo.

En este anexo se describen los proyectos que están en curso, completados o anunciados por los miembros de IVOA. Debido a que cada miembro es libre de tener o no anunciados sus proyectos relacionados al OV, no es fácil encontrar una descripción de cada proyecto, es por ello que es posible que no se encuentren descritos todos los proyectos en este listado.

C.1. Argentina Virtual Observatory

NOVA fue fundado en Enero del 2009 por 8 importantes instituciones astronómicas y la facultad de Ciencias astronómicas y geofísicas de la Universidad Nacional de la Plata. Desde Junio del año 2013 NOVA espera comenzar sus operaciones proveyendo observaciones astronómicas desde su sitio oficial e implementar una plataforma web donde se pueda trabajar con los datos.

Proyectos

- NOVA@IAFE: Construirá una base de datos para observaciones realizadas por el telescopio solar HASTA y sus aplicaciones.
- NOVA@ICATE: Construirá una base de datos para las observaciones espectroscópicas disponibles en ICATE.

C.2. Armenian Virtual Observatory

ArVo está basado en el *Digital First Byukaran Survey* (DFBS), proyecto realizado en conjunto por el observatorio astrofísico Byukaran ubicado en Armenia; *Università di Roma “La Sapienza”*, Italia; Cornell University, USA; y VO-Franche, Francia. Este OV fue lanzado en Febrero del 2008.

Proyectos

- Búsqueda de nuevos objetos interesantes con ciertos tipos de patrones de baja dispersión espectral.
- Identificación vía óptica de nuevas fuentes gamma, rayos X, Infrarrojos y de radio.

- Identificación de las nuevas fuentes infrarrojas encontradas en el Telescopio Espacial Spitzer.

C.3. AstroGrid

AstroGrid es el OV que pertenece al Reino Unido. Comenzó el año 2001 como proyecto y en Abril del 2008 fue lanzado junto con servicios funcionales y *softwares* para usuarios. Ha sido financiado por el Consejo de Investigación de Partículas Físicas y Astronomía (PPARC, por sus siglas en inglés) y el Consejo de Instalaciones Científicas y Tecnológicas (STFC, por sus siglas en inglés).

Proyectos

- Topcat: Software interactivo que permite visualizar gráficamente y editar datos tabulados en formatos como FITS y VOTable.
- VODesktop: Conjunto de herramientas con tecnología de OV.
- AstroRuntime: Provee servicios de OV a VODesktop y otras aplicaciones, simplificando el uso de estos servicios desde el código de las aplicaciones.

C.4. Australian Virtual Observatory

El trabajo en el proyecto AusVO comenzó en el 2004, financiado parcialmente por el Consejo Australiano de Investigación de Infraestructura de Conectividad y con una gran variedad de instituciones asociadas entre Universidades, Telescopios y Observatorios australianos. Es miembro fundador del IVOA desde Junio del 2002, mientras que los astrónomos australianos han estado activos en el ecosistema OV desde Diciembre del 2000.

Proyectos

- RAVE: Empresa multinacional que busca medir las velocidades radiales y metalicidades de una gran cantidad de estrellas en el vecindario del sistema solar.
- *Remote Visualisation System (RVS)*: *Software* que permite visualización remota y análisis de conjuntos de datos astronómicos. Actualmente, en fase de prototipo.
- *Distributed Volume Renderer (dvr)*: *Software* diseñado para renderizar conjuntos de datos volumétricos de tamaño mayores a la memoria del hardware donde corre el

software con altas velocidades de respuesta ante la interacción del usuario usando clústeres tipo *Beowulf*.

- 2QZ: Sondeo astronómico con cerca de 22000 QSO identificados espectroscópicamente utilizando un instrumento de un campo de 2 grados ubicado en el *Anglo-Australian Telescope* (AAT).
- *Machine Learning for Source Matching*: Proyecto enfocado en combinar y encontrar los mismo objetos en distintas fuentes. Planean utilizar el catálogo HIPASS y el *SuperCOSOS Sky Surveys* (SSS).
- *Massive Compact Halo Objects* (MACHO) *Archive*: Datos obtenidos entre 1992 y 1999 que actualmente son accesibles a través de una interfaz web y se encuentra en proceso de ser llevado al OV.
- *Australia Telescope Compact Array* (ATCA) *Archive*: Radiotelescopio de apertura sintética que generó 1.5 TeraBytes, actualmente se encuentra disponible online.
- *HI Parkes All Sky Survey* (HIPASS): *Survey* de Hidrógeno atómico neutral realizado entre 1997 y 2002, actualmente disponible a través del OV.

C.5. Brazilian Virtual Observatory

La intención de realizar el BRAVO nació en Abril del 2008 con 6 institutos brasileños de investigación y la Sociedad Astronómica Brasileña (SBA, por sus siglas en portugués). Luego fue fundado por el Instituto Nacional de Ciencias y Tecnologías en Astrofísicas (INCT-A).

Proyectos

- BRAVO@INPE: Generar inversión en Tecnologías de la información en infraestructura computacional, *Data Grid*, procesamiento de datos y Minería de datos.
- BRAVO@LNA: OV del *Southern Astrophysical Research Telescope* (SOAR) para los astrónomos brasileños.
- BRAVO@UFSC: Investigación del poder de síntesis espectral para estimar las propiedades físicas de las galaxias.

- CYCLOPS: *Software* que modela las emisiones ópticas de sistemas AM Her, incluyendo los cuatro parámetros de Stokes.

C.6. Canadian Virtual Observatory

Observatorio Virtual Canadiense creado por el Centro de Datos Astronómicos Canadiense (CADC, por sus siglas en inglés).

Proyectos

- *Data Sharing* (VOspace 2.0): Servicio que permite a los usuarios compartir archivos y colaborar con miembros de sus equipos.
- *Table Access Protocol* (TAP-1.0): Servicio que permite el acceso a los datos descritos por el *Common Archive Observation Model* (CAOM) utilizado por el CADC, así como tablas de otros proyectos.
- *Observation Model Core Components* (ObsCore-1.0): Modelo que implementa una vista estándar para el TAP-1.0.
- *Simple Image Access* (SIA-1.0): Servicio compatible con el estándar SIA-1.0 de IVOA, permite realizar consultas para fácil acceso de las imágenes calibradas de la mayoría de las colecciones de datos del CADC.

C.7. Chinese Virtual Observatory

El China-VO inició en el 2002 por la Academia China de Ciencias, Observatorios Astronómicos Nacionales, 4 socios y 12 colaboradores.

Proyectos

- *FITS Manager* (FM): Aplicación descargable que permite gestionar archivos FITS, VOTable, entre otros, que están almacenados en computadores personales. Es un proyecto en colaboración con VOI.
- *VO Data Access Service* (VO-DAS): Framework para acceso a datos que permite consultar grandes volúmenes de recursos astronómicos como catálogos, espectros e imágenes a través de líneas de comando, interfaz gráfica o un sitio web.
- *FITS Header Archiving System* (FitHAS): Herramienta descargable que permite ver e importar las cabeceras de uno o varios archivos FITS a una tabla de una base de datos. También fue desarrollado por el Centro IBM, Universidad Tianjin (TU) y el

centro de investigación de aplicación e-science del Centro de información de redes en la computación (CNIC, por sus siglas en inglés) de la Academia China de las Ciencias (CAS, por sus siglas en inglés).

– *Imaging Processing and analysis tool for China VO* (VO IMPAT): Herramienta en JAVA descargable para procesar y analizar imágenes, permite visualizar imágenes del cielo y acceder a los datos relacionados del Centro de Datos Astronómicos de Beijing (BADC, por sus siglas en inglés).

– VOTable2XHTML: Un stylesheet XSL (*EXtensible Stylesheet Language*) que puede ser utilizada para transformar archivos en VOTable en archivos XHTML.

– SkyMouse: Motor de búsquedas que permite acceder servicios astronómicos como servicios web CGI (*Common Gateway Interface*).

C.8. European Space Agency

Miembro activo y co-fundador del proyecto EURO-VO. Su objetivo es asegurarse que todos los archivos de (*European Space Astronomy Centre*) ESAC sean accesibles a través de tecnologías de OV, crear aplicaciones y herramientas para el OV, educar sobre el OV, construir un GRID computacional entre otros.

Proyectos

– DALToolKit: *Software* descargable en JAVA que permite publicar en los OV a través del protocolo de la capa de acceso a datos de IVOA (DAL, por sus siglas en inglés). Convierte consultas en DAL a consultas específicas a una base de datos en SQL, luego serializa los resultados de la consulta en VOTables.

– *IVOA Resource Registry*: El registro de recurso de IVOA oficial para el EURO-VO.

– VOSpec: Herramienta de análisis espectral de múltiples longitudes de con acceso a bases de datos de átomos y moléculas y sus modelos espectrales y teóricos registrados en el OV. Puede ser descargada o utilizada a través del Internet.

– *Science Activities in the VO*: Desarrollo de proyectos de investigación basados en el OV, tutoriales que enseñan el uso de herramientas realizadas por el Equipo de

Archivos Científicos (SAT, por sus siglas en inglés), entre otros. Desarrollados por el Centro de Astronomía Espacial Europea (ESAC, por sus siglas en inglés).

C.9. European Virtual Observatory

El EURO-VO es la continuación del Astrophysical Virtual Observatory (AVO), proyecto de la European Commission y 6 organizaciones encargado de investigar de los requerimientos y tecnologías necesarias para construir el European Virtual Observatory. El Objetivo principal del EURO-VO es implementar y hacer funcionar el OV en Europa.

Proyectos

- VOTECH: Primer proyecto que implementó el concepto del Centro de Tecnologías de EURO-VO (EURO-VOTC) como parte del EURO-VO.
- EURO-VO *Data Centre Alliance* (EuroVO-DCA): Respaldado por la Unión Europea en el contexto de la iniciativa “FP6 e-Infrastructure Communication Network Development”.
- EURO-VO *Astronomical Infrastructure for Data Access* (EuroVO-AIDA): Respaldado por la Unión Europea en el contexto de la iniciativa “FP7 e-Infrastructure Scientific Research Repositories”
- Euro-VO *International Cooperation Empowerment* (EuroVO-ICE): Respaldado por la Unión Europea en el contexto de la iniciativa “FP7 INFRA-2010-2.3.3 Research Infrastructures”.
- EuroVO-CoSADIE: Respaldado por la Unión Europea en el contexto de la iniciativa “FP7 INFRA-2012-3.3 Research Infrastructure”.

C.10. German Astrophysical Virtual Observatory

GAVO fue lanzado en el 2003. Es financiado a través del *Federal Ministry of Education and Research* (BMBF).

Proyectos

- GAVO *Data Center*: Colección de datos y servicios en crecimiento provistos por terceros y GAVO.

- *MPA Simulations access*: Servicio Web para consultar vía SQL los resultados de la simulación *Millennium*.
- *MultiDark Database*: Servicio que da acceso a los datos de las simulaciones MultiDark y Bolshoi, a través de consultas SQL. Basado en la aplicación Web de *Millennium*.
- *RAVE archive search*: Acceso al creciente archivo de velocidades radiales de más de 400.000 estrellas.
- *TheoSSA*: Servicio que provee distribuciones de energía espectrales basado en cálculos de modelos atmosféricos.

C.11. Hungarian Virtual Observatory

Proyecto que busca coordinar los esfuerzos de los investigadores húngaros que trabajan en el desarrollo de OV's.

Proyectos

- *Spectrum Service for VO*: Propuesta cuyo propósito es agregar diversas características y realizar mejoras a los servicios web que contienen galaxias espectrales y otros objetos astronómicos.
- *Synthetic Spectrum Service*: Propuesta con el propósito de ofrecer como servicio web los datos espectrales.
- *Photometric Redshift*: Propuesta con el objetivo de realizar un servicio web que estime el corrimiento al rojo de fotometrías incrementales por dos órdenes de magnitud la cantidad de objetos a una distancia conocida.
- *Linking Web Services to GRID clusters*: Propuesta que espera mejorar la comunicación entre Sistemas Operativos debido a que las simulaciones por lo general se encuentran optimizadas para un solo Sistema Operativo y generalmente, es infactible reescribir el código de la simulación para otro Sistema Operativo.
- *Information Bulletin on Variable Stars*: Boletín en nombre de las Comisiones 27 y 42 de la Unión Astronómica Internacional, publicado por el Observatorio Konkoly de la Academia de Ciencias Húngaro.

- *Debrecen Photoheliographic Data* (DPD): Catálogo de manchas solares con posiciones heliográficas y sus correspondientes áreas.

C.12. Italian Virtual Observatory

El VObs.it es miembro de IVOA y EURO-VO. Fundado y llevado a cabo por el Instituto Nacional Italiano de Astrofísica (INAF, por sus siglas en italiano).

Proyectos

- SIAP: Servicio que provee acceso al HST-ACS GOODS dentro del VVDS-CDFS.
- SSAP: Permite el acceso al VVDS-F02-DEEP a través de tecnologías OV. Contiene datos espectrales de 8980 objetos.
- CONE SEARCH: Permite realizar consultas al VVDS-CDFS.
- SKYNODE: Permite realizar consultas a los catálogos del VVDS.

C.13. Japanese Virtual Observatory

El JVO fue implementado por el *National Astronomical Observatory of Japan* (NAOJ) en colaboración con Fujitsu para el desarrollo de los prototipos de los sistemas del JVO. Inició en el 2002 y sus servicios están disponibles desde el 2004, un sistema de análisis fue integrado en el 2005.

Proyectos

- *JVO portal service*: Portal Web para el acceso de distintas fuentes como el Telescopio Subarú, *Sloan Digital Sky* y ALMA, entre otros.

C.14. Observatoire Virtual France

No hay información detallada de sus proyectos, pero se encuentra activo a través de workshops y asistencia a reuniones de IVOA.

C.15. Russian Virtual Observatory

Proyecto que fue realizado debido a que la mayoría de los observatorios que se encontraban en la Ex Unión Soviética, ahora se encuentran fuera de Rusia, fundado por el Consejo de Investigación Astronómica (RAS, por sus siglas en inglés)

Proyectos

- *Systems for Knowledge Integration* (SYNTHESIS): *Framework* desarrollado por el grupo del mismo nombre
- INFOSEM: Su objetivo principal es investigar, prototipar y diseminar metodologías y técnicas básicas, permitiendo la construcción de sistemas de información interoperables entre fuentes de información heterogéneas.
- SEMIMOD: Modelado y administración de datos semi estructurados para aplicaciones web dinámicas.
- BIOMED: Métodos y herramientas para el desarrollo de mediadores para librerías digitales distribuidas.
- VOINFRA: Desarrollo de la base interoperable de la infraestructura del RVO.
- MULTISOURCE: Métodos para resolver problemas heterogéneos de distintas fuentes de información.
- ASTROMEDIA: Métodos y herramientas para trabajar con la arquitectura AstroGrid para el RVO.
- ConcMod: Desarrollo de herramientas para la definición de modelos conceptuales y resolución de problemas en sistemas GRID.
- RuleInt: Integración de programas basados en reglas y bases de datos de conocimientos y servicios para resolución de problemas científicos en fuentes de información heterogéneas y distribuidas.
- ASTROMEDIA Trial: Arquitectura híbrida del AstroGrid y Mediator *Middleware*.
- *Galaxies Search*: Búsqueda de galaxias distantes sobre AstroGrid.
- *Star Classification*: Clasificación de estrellas binarias eclípticas con el algoritmo *Ensembled* Weka sobre AstroGrid.

C.16. Spanish Virtual Observatory

El SVO inició en Junio del 2004, los miembros son el centro de Astrobiología (INTA-CSIC), el Departamento de inteligencia Artificial de la Universidad Nacional de Educación Distante (UNED), Universidad de Cádiz y el Centro de Servicios Científicos y Académicos de Cataluña (CESCA).

Proyectos

- VOSA: Herramienta que permite analizar datos estelares y galácticos a través de la lectura de tablas fotométricas del usuario. Permite consultar múltiples catálogos fotométricos accesibles a través de servicios de OV, modelos teóricos espectrales compatibles con los OV, realizar pruebas estadísticas para determinar cuál modelo reproduce mejor los datos observados, entre otros.
- VOSED: Servicio que construye Distribuciones de Energía Espectrales al obtener información de servicios espectroscópicos en los OV.
- TESELA: Servicio que permite acceder a catálogos de regiones vacías. Está basado en la aplicación de la triangulación de Delaunay sobre el cielo.

C.17. US Virtual Astronomical Observatory

El VAO es el sucesor del *National Virtual Observatory* (NVO) y fue fundado por el NSD y la NASA. Está a cargo de VAO LLC, *Associated Universities Inc.* (AUI) y *Association of Universities of Research in Astronomy* (AURA). El US VO es co-fundador de IVOA.

Proyectos

- *Data Discovery Tool*: Herramienta web que permite encontrar conjuntos de datos de colecciones astronómicas conocidas para el OV como el Telescopio Espacial Hubble, el Observatorio Chandra de Rayos-X, el Telescopio Espacial Spitzer, entre otros.
- *Iris, SED Analysis Tool*: Aplicación descargable para encontrar, graficar y corregir las Distribuciones de Energía Espectrales.
- *Cross-Comparison Tool*: Herramienta web que realiza comparaciones cruzadas entre una tabla provista por el usuario y otras disponibles en línea del catálogo de fuentes, dentro de un radio especificado por el usuario. Retorna todas las fuentes encontradas en el catálogo dentro del radio.
- *Time Series Search Tool*: Herramienta web que provee acceso a datos de series de tiempo del Centro de Series de Tiempo de Harvard, del Archivo de Exoplanetas de la NASA y del Sondeo de eventos Transitorios en Tiempo Real de Catalina, y los

analiza dentro de la aplicación de periodograma del Archivo de Exoplanetas de la NASA.

C.18. Virtual Observatory India

El VOI es una colaboración entre el *Inter University center for Astronomy and Astrophysics* (IUCAA) y *Persistent System, Ltd.*, y apoyado por el Ministerio de Comunicaciones y Tecnologías de la Información.

Proyectos

- VOIPortal: Entrada a todos los servicios web de VOI. Es posible navegar los datos descargándolos o a través de las aplicaciones web VOIMosaic y PyMorph.
- *Mosaic Service: Software* que permite realizar mosaicos con SWarp, SExtractor y STIFF de las imágenes del SDSS, 2MASS y HST.
- *PyMorph Service: Software* que permite derivar parámetros morfológicos a partir de imágenes de galaxias. Puede recibir como entrada los archivos FITS generados por el Mosaic Service.
- VOPlot: Herramienta de *Software* en JAVA que permite visualizar datos astronómicos disponibles en VOTable, ASCII y FITS.
- VOMegaPlot: Herramienta de *Software* desarrollada en JAVA que permite visualizar datos astronómicos disponibles en VOTable, al igual que VOPlot. Tiene una versión cliente-servidor.
- AstroStat: Herramienta de *Software* que permite a los astrónomos utilizar rutinas estadísticas on grandes conjuntos de datos subidos en VOTable o ASCII.
- VOCat: Herramienta de *Software* que convierte los catálogos astronómicos en bases de datos MySQL.
- VOPlatform: Software desarrollado en JAVA que permite a los usuarios organizar sus herramientas de OV, conjuntos de datos y otros recursos como documentos, enlaces y otros en un solo lugar.
- VOConvert (ConVOT): *Software* que convierte ASCII a VOTables, FITS a VOTable. y VOTable a ASCII.

- *Android Cosmological Calculator*: Aplicación Android que permite a los usuarios obtener la edad actual del universo, la distancia, movimiento radial, volumen, el tamaño angular de la distancia, corrimiento al rojo y la distancia lumínica en base a parámetros establecidos por el usuario,
- *Android Name Resolver*: Aplicación Android que, al introducir el nombre de un objeto celestial, retorna información respecto a éste como valores RA/Dec, corrimiento al rojo, entre otros.
- *Fits Manager*: Herramienta Web para visualizar, crear y agregar extensiones y convertir archivos FITS.
- *HCT Data Archive System*: Sistema web que archiva los datos generados por el *Himalayan Chandra Telescope* (HCT).