



FONDEF R&D "D11i1060"

Development of an Astro-Informatic Platform for Management and Intelligent Analysis of Large-scale Data

Mauricio Solar

msolar@inf.utfsm.cl

Universidad Técnica Federico Santa María (UTFSM)

<http://youtu.be/sVNLU8wVjIQ>

2013 – Santiago, Chile

Agenda



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Introducción

- Astronomía en Chile

Astro-Ingeniería

- Áreas de la Astronomía

Técnicas Computacionales en Astronomía

- Data Mining, Machine Learning and HPC

Proyecto FONDEF de Astro-Informática

- IVOA, ChiVO, Team, Sponsors, etc.

Agenda



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Introducción

- Astronomía en Chile

Astro-Ingeniería

- Áreas de la Astronomía

Técnicas Computacionales en Astronomía

- Data Mining, Machine Learning and HPC

Proyecto FONDEF de Astro-Informática

- IVOA, ChiVO, Team, Sponsors, etc.

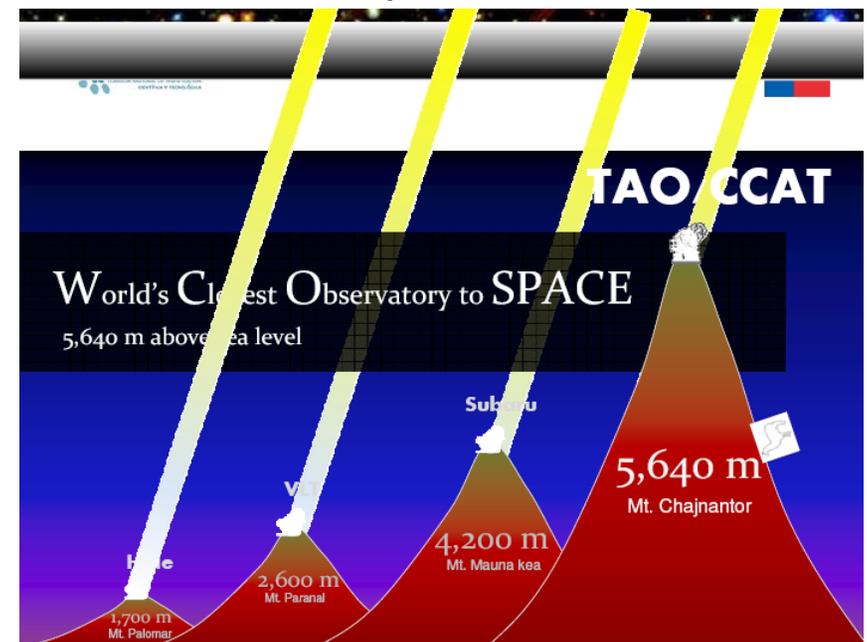
Introducción



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Cielos chilenos tienen condiciones atmosféricas únicas para la observación astronómica:

- Más de 330 noches de cielo despejado al año.
- Llano de **Chajnantor** (lengua kunza: "lugar de partida/despegar"):
- lugar de la tierra con las mejores condiciones climáticas para observar el *espectro milimétrico*
- es el *mas alto del mundo*.



Introducción



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

- Principales instituciones de investigación astronómica en Chile:
 - *Association of Universities for the Research in Astronomy* (**AURA**)
 - *European Southern Observatory* (**ESO**)
 - *The National Radio Astronomy Observatory* (**NRAO**)
 - *National Astronomical Observatory of Japan* (**NAOJ**)
 - *Carnegie Institution for Science*.
- **Chile:** mayor concentración instrumental astronómica del mundo:
 - Más de 30 observatorios (6 de proyectos internacionales).
 - Año 2018 albergará casi 70% de capacidad de observación astronómica mundial (potencia mundial en actividad astronómica).
- **Expertos de NU y NSF:** la astronomía es una de las ciencias que ha alcanzado mayor nivel en Chile (comparable a países desarrollados).

Evolución de inversión en Chile



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

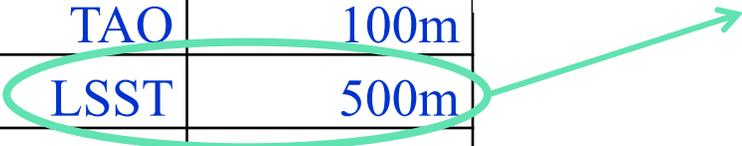
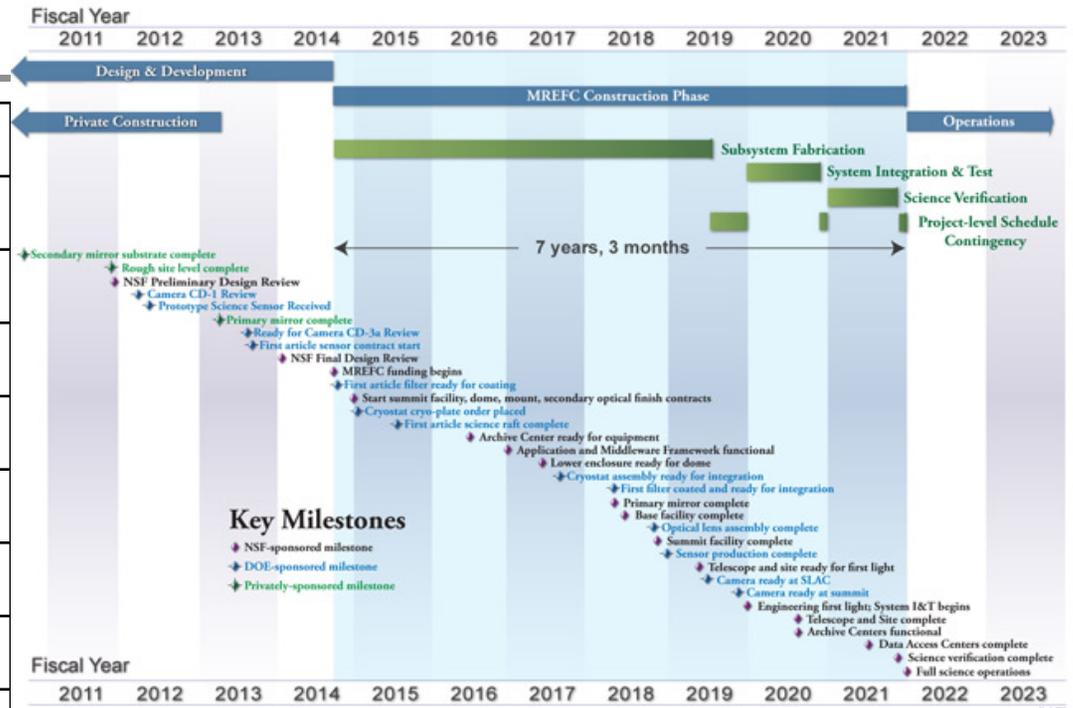
En operación	Costo USD
Telescopio Magallanes	100m
Gemini	300m
VLT	700m
ACT	40m
En construcción	
ALMA	1,300m
Proyectado	
TAO	100m
LSST	500m
CCAT	200m
GMT	800m
E-ELT	1,500m
Total	6,100m

Evolución de inversión en Chile



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

En operación	Costo USD
Telescopio Magallanes	100m
Gemini	300m
VLT	700m
ACT	40m
En construcción	
ALMA	1,300m
Proyectado	
TAO	100m
LSST	500m
CCAT	200m
GMT	800m
E-ELT	1,500m
Total	6,100m

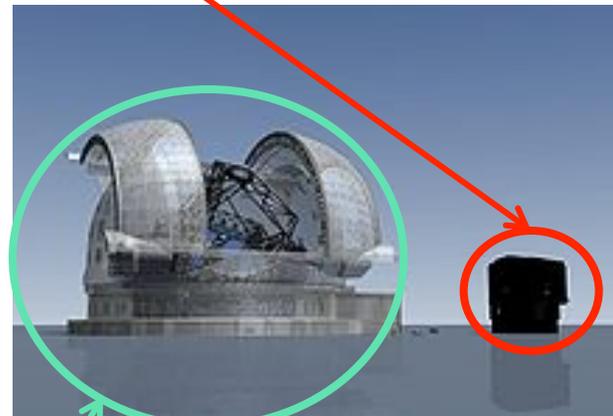


Evolución de inversión en Chile



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

En operación	Costo USD
Telescopio Magallanes	100m
Gemini	300m
VLT	700m
ACT	40m
En construcción	
ALMA	1,300m
Proyectado	
TAO	100m
LSST	500m
CCAT	200m
GMT	800m
E-ELT	1,500m
Total	6,100m

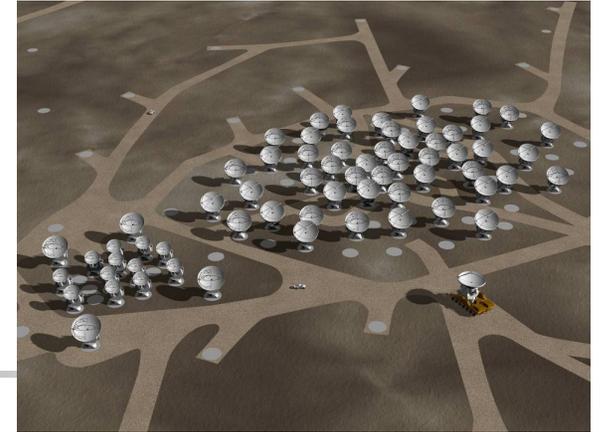


Introduction



- The **Atacama Large Millimeter Array (ALMA)**:
 - Site: Chajnantor plateau (Chile) @5000 masl
 - Worldwide collaboration: Europe (ESO), USA (NRAO), East Asia (NAOJ)
 - USD 1.3 Billions, most expensive ground based telescope.
- It is the biggest radio-telescope with 66 antennas, distributed over a wide extension, with up to 16 km of separation.
- ALMA will provide the possibility to be used as a single array, or as up to six minor independent arrays or groups of antennas:
 - Each array is equivalent to one instrument
 - Can be seen as a multi-telescope problem.

Introduction



- The antennas:
 - will be changing their position during the year,
 - different distributions will be used to exploit different observations.
- One of the major goals of automatic scheduling is (ECOS-Conicyt):
 - to optimize the observations efficiency,
 - by maximizing the use of the expensive available operation time, and
 - schedule of required maintenance.
- ALMA (radio-telescope) observations:
 - not limited to nighttime
 - a 24/7 operation
 - ALMA **will generate 1-5 TB** of data by week





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Agenda

Introducción

- Astronomía en Chile

Astro-Ingeniería

- Áreas de la Astronomía

Técnicas Computacionales en Astronomía

- Data Mining, Machine Learning and HPC

Proyecto FONDEF de Astro-Informática

- IVOA, ChiVO, Team, Sponsors, etc.

Astro-ingeniería



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astroingeniería: la "*ingeniería de instrumentos astronómicos*".

Diversidad de Instrumentos astronómicos depende de su función:

Colección de fotones desde fuentes astronómicas: óptica (sistema de reflexión y refracción): espejos, lentes, paneles reflectores, tecnología de revestimiento y difusión, etc.

Detección de fotones: cámaras CCD (*Charge-Coupled Device*), tecnología de lectura de CCD, fotómetros, detectores de potencia, diodos, sensores de bolómetros, etc.

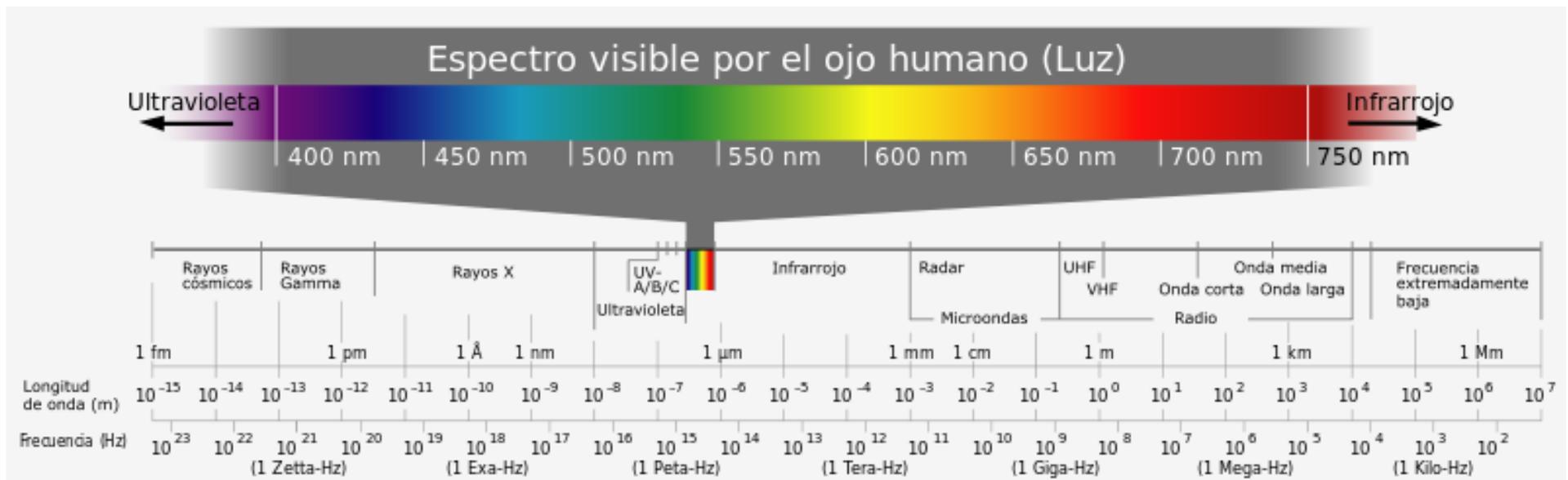
Tecnología auxiliar: crióstatos, sensores (detectores ambientales, de posición y movimiento), sistemas de posicionamiento de espejos opto-mecánicos (sistemas cuadrúpedo, hexápodos, etapas de traducción mecánicas), óptica activa y adaptativa, lasers, etc.

Astro-Informática: Procesamiento de datos, imágenes y espectroscopía. Sistemas complejos mecano-electro-ópticos para procesar la luz detectada, software de procesamiento de datos.

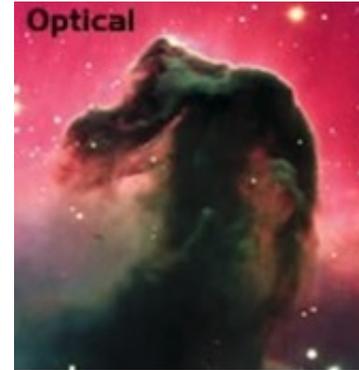


Astro-ingeniería

Astronomía observacional clásica - *espectro electromagnético*

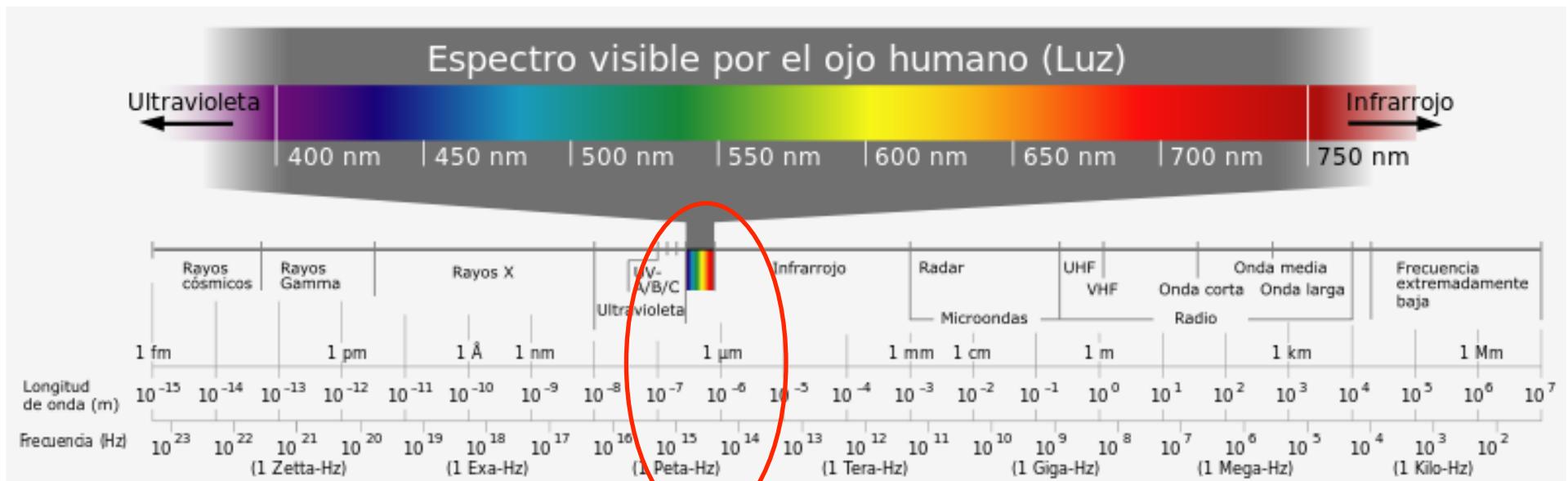


Astro-ingeniería



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astronomía observacional clásica - *espectro electromagnético*



Astronomía óptica: usa componentes ópticos (espejos, lentes, detectores de estado sólido) para observar λ cercana al Infrarrojo (**IR**) hasta la cercana al Ultravioleta (**UV**).

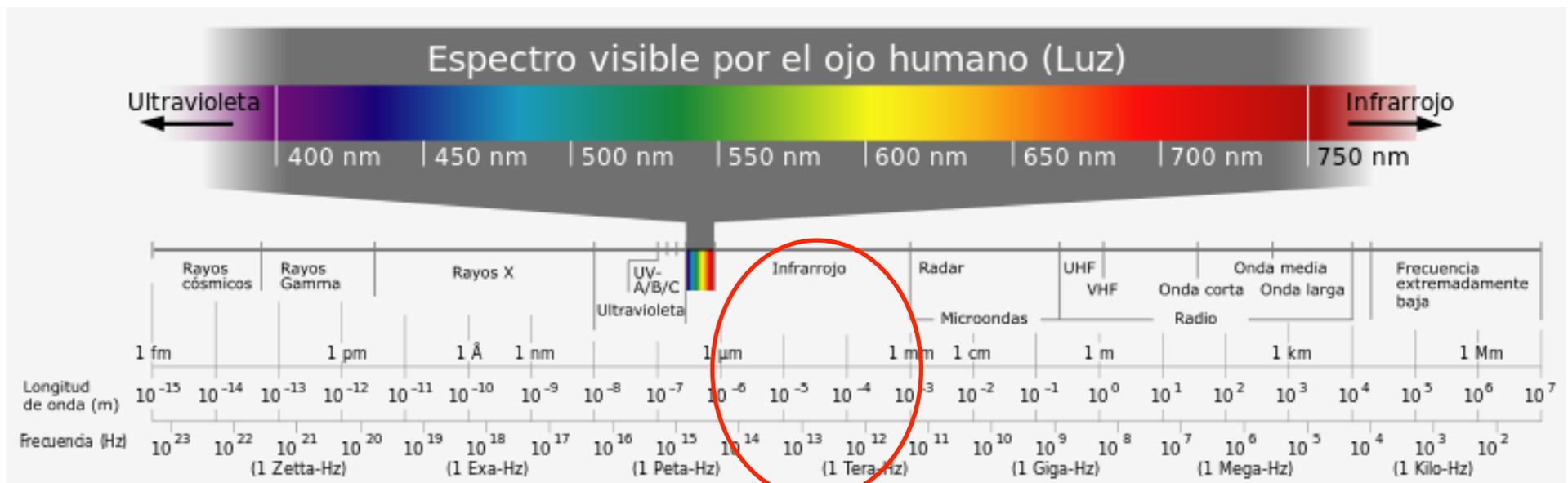
Astronomía de luz visible (λ detectadas al ojo 400-700nm) en medio del rango

Astro-ingeniería



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Astronomía observacional clásica - *espectro electromagnético*

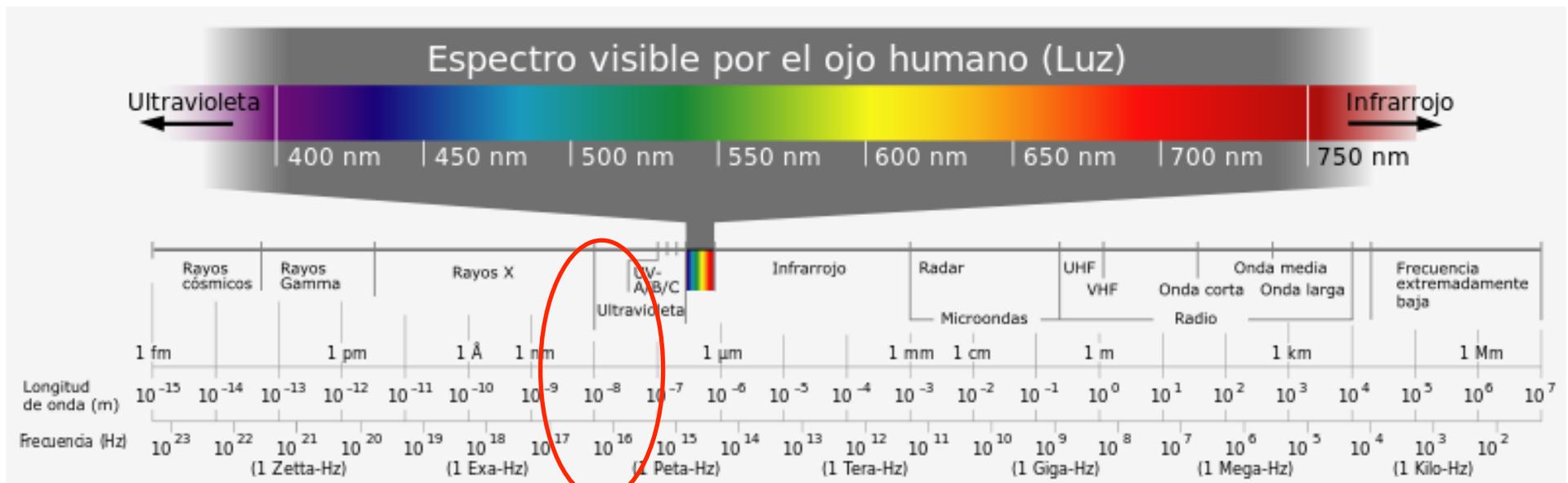


Astronomía de IR: detección y análisis de radiación IR (se refiere a λ más largas que el límite de detección de los detectores de estado sólido de silicio, $\lambda \approx 1 \mu\text{m}$). Instrumento más común: telescopio reflector, pero con detector sensible a λ IR. Telescopios espaciales se utilizan en ciertas λ donde la atmósfera es opaca, o para eliminar el ruido (radiación térmica de la atmósfera).



Astro-ingeniería

Astronomía observacional clásica - *espectro electromagnético*



Astronomía de alta energía incluye:

astronomía UV extrema

astronomía de rayos X

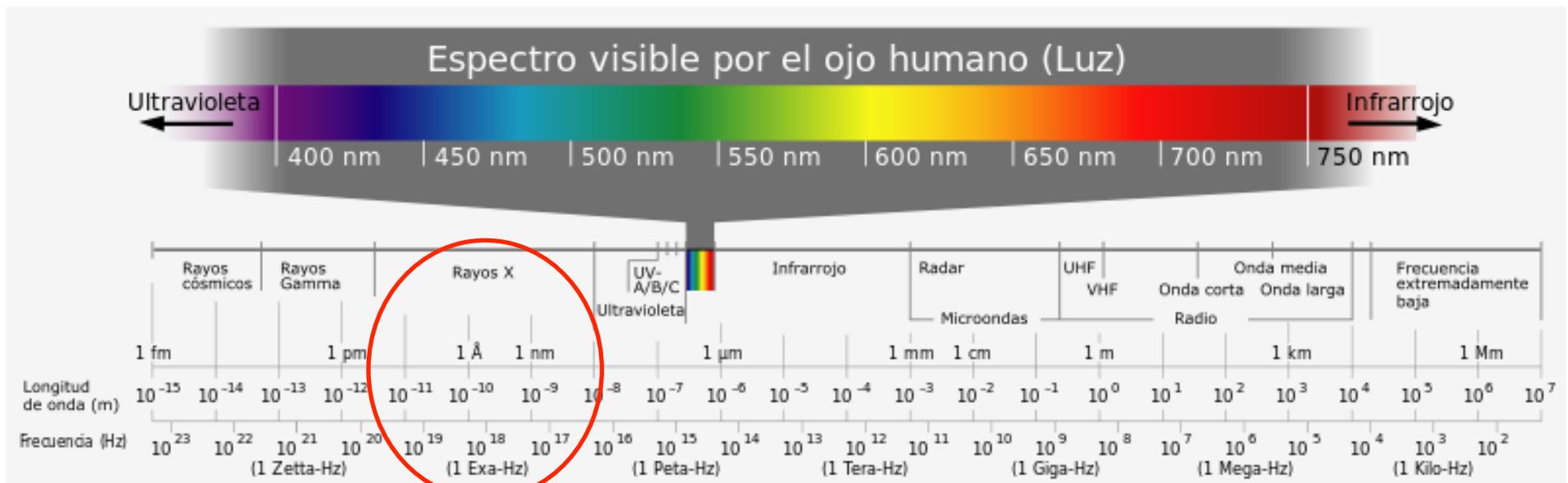
astronomía de rayos gamma (γ)

estudios de neutrinos y rayos cósmicos.



Astro-ingeniería

Astronomía observacional clásica - *espectro electromagnético*

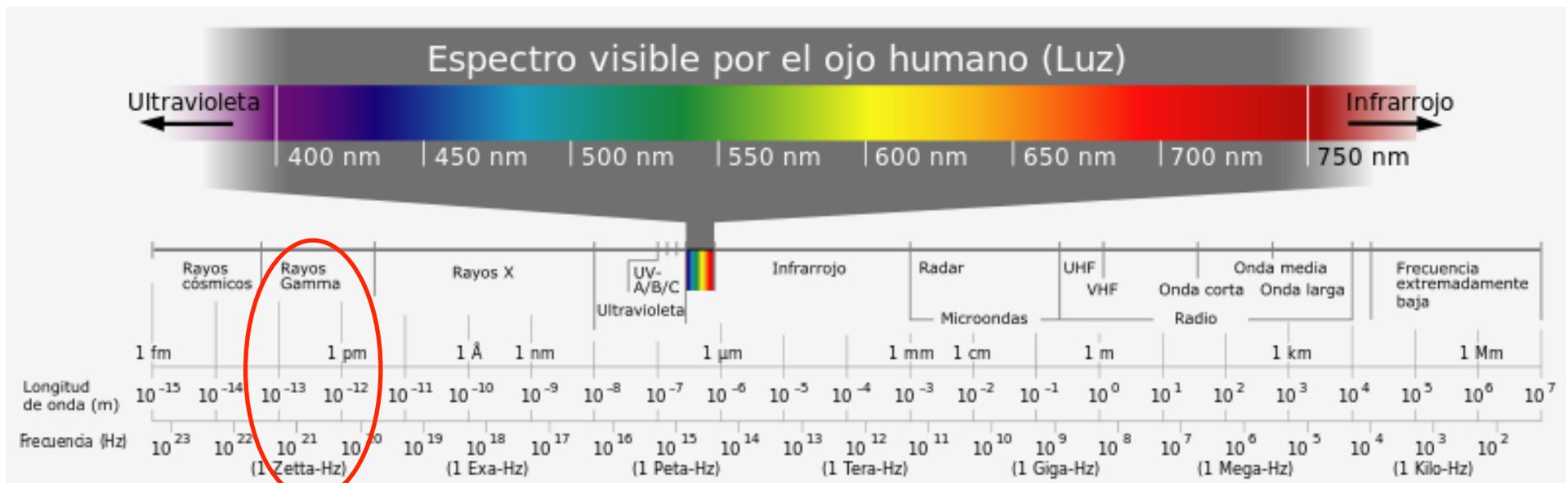


Astronomía de alta energía incluye:
astronomía UV extrema
astronomía de rayos X
astronomía de rayos gamma (γ)
estudios de neutrinos y rayos cósmicos.



Astro-ingeniería

Astronomía observacional clásica - *espectro electromagnético*

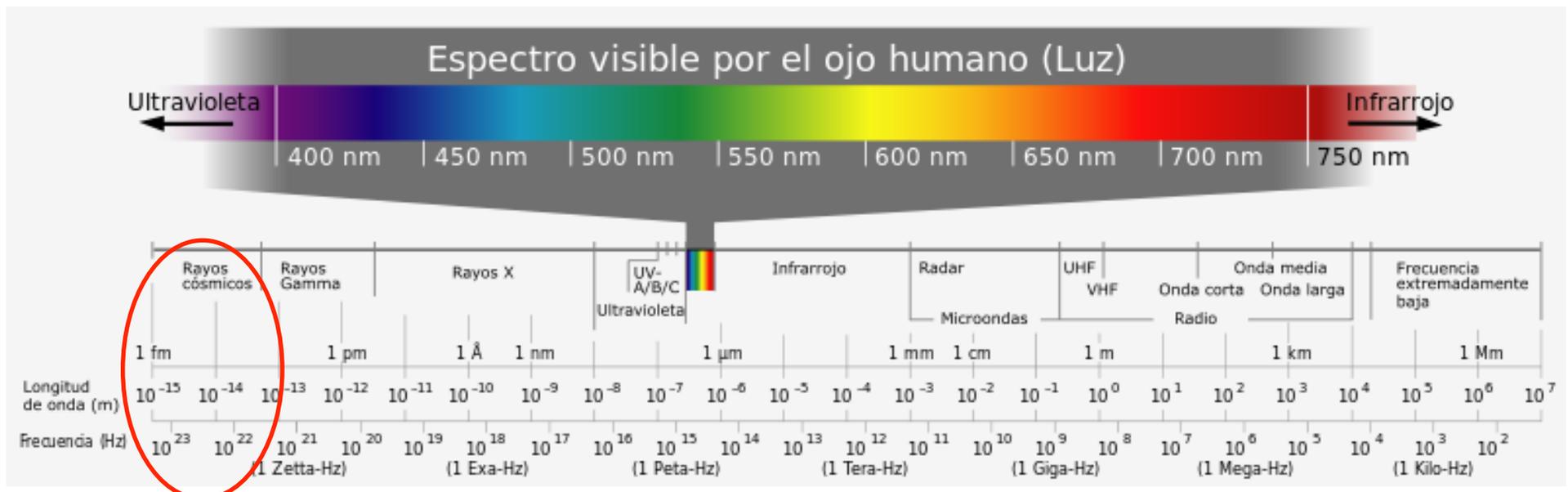


Astronomía de alta energía incluye:
astronomía UV extrema
astronomía de rayos X
astronomía de rayos gamma (γ)
estudios de neutrinos y rayos cósmicos.



Astro-ingeniería

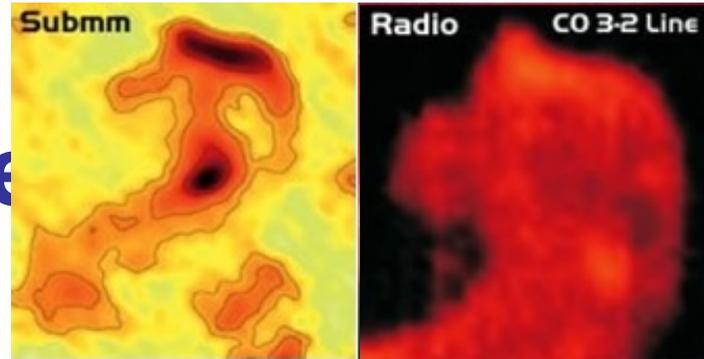
Astronomía observacional clásica - *espectro electromagnético*



Astronomía de alta energía incluye:

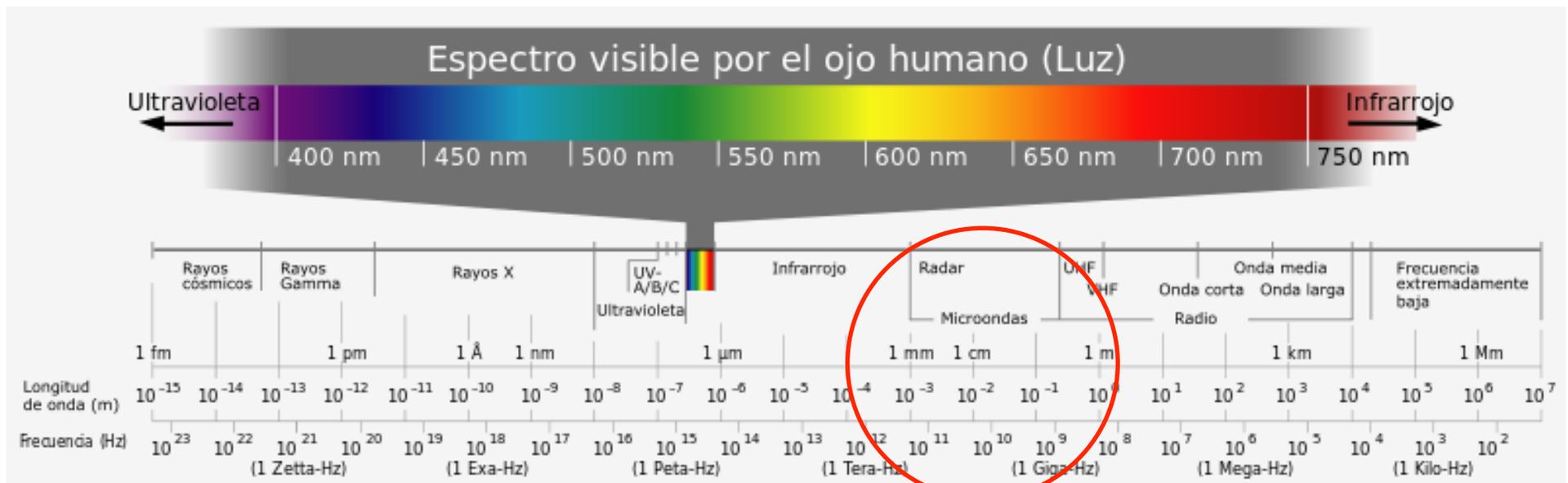
- astronomía UV extrema**
- astronomía de rayos X**
- astronomía de rayos gamma (γ)**
- estudios de neutrinos y rayos cósmicos**

Astro-ingeniería

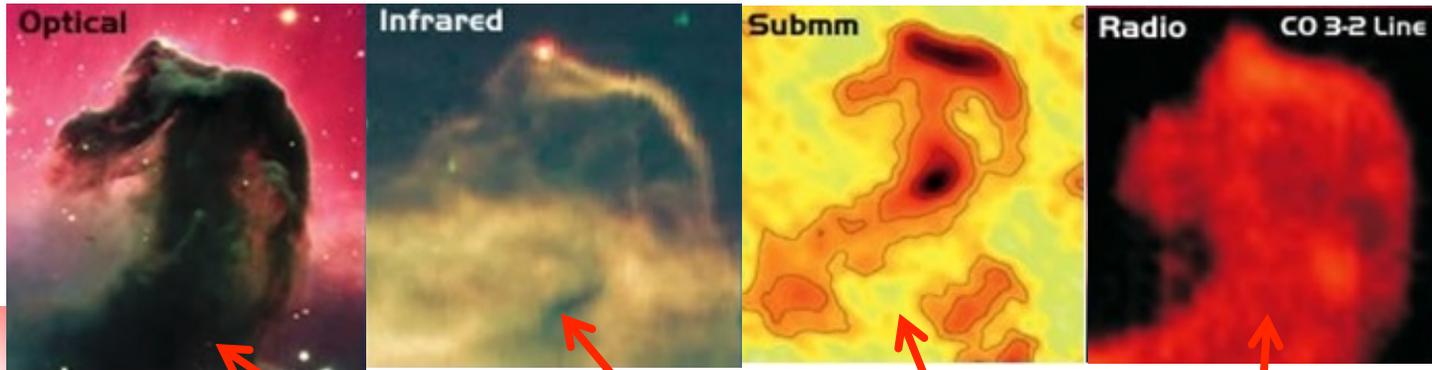


UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astronomía observacional clásica - *espectro electromagnético*

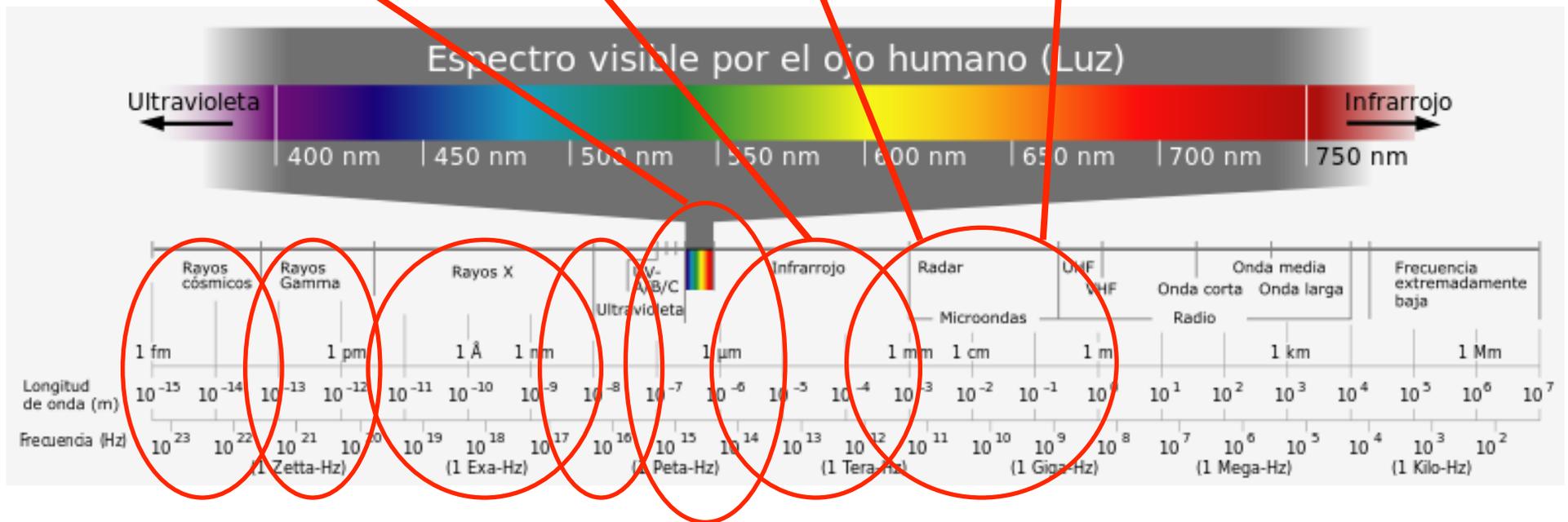


Radioastronomía: detecta radiación de λ desde 1 mm hasta λ decamétricas. Receptores similares a los de transmisión de radiodifusión, pero mucho más sensible. También hay telescopios de radio.



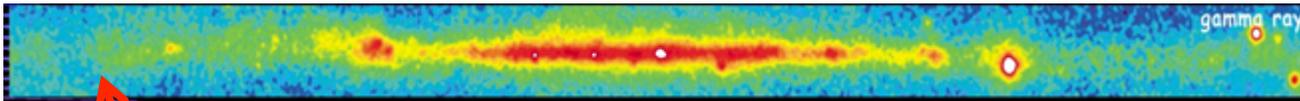
UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astronomía observacional clásica - *espectro electromagnético*



Instrumentos: extienden nuestra "vista" en las λ del espectro electromagnético que permite saber más sobre la estructura y dinámica del universo.

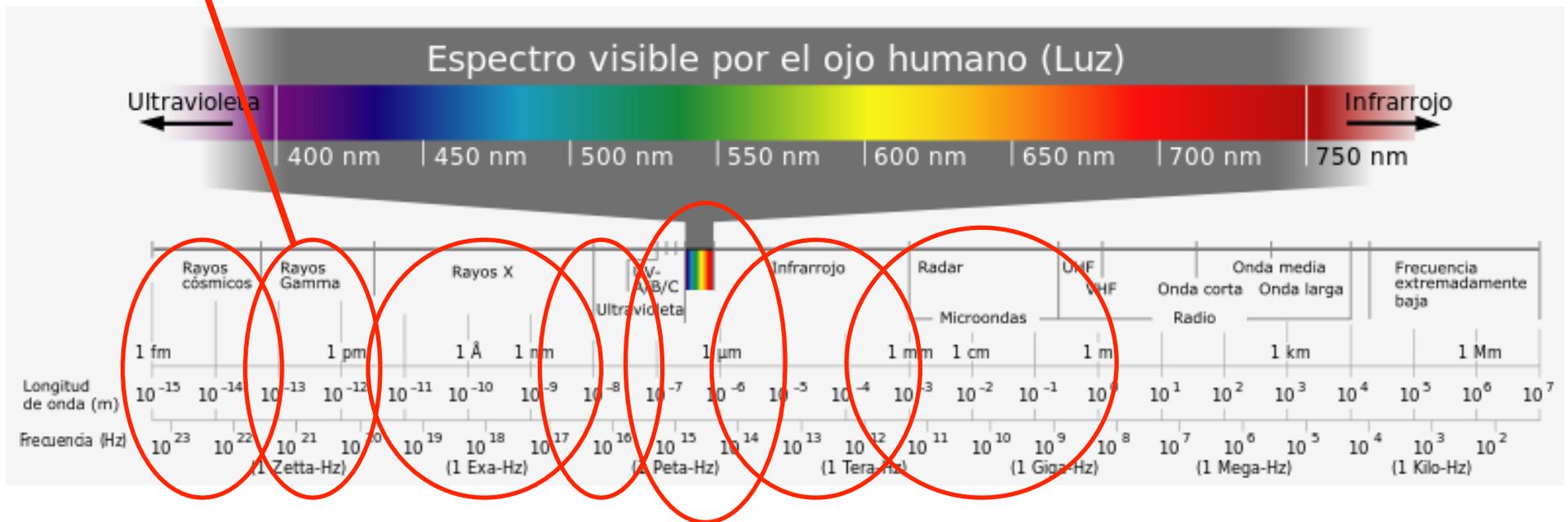
Astronomía de **múltiples** λ revela el universo en toda su luz.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

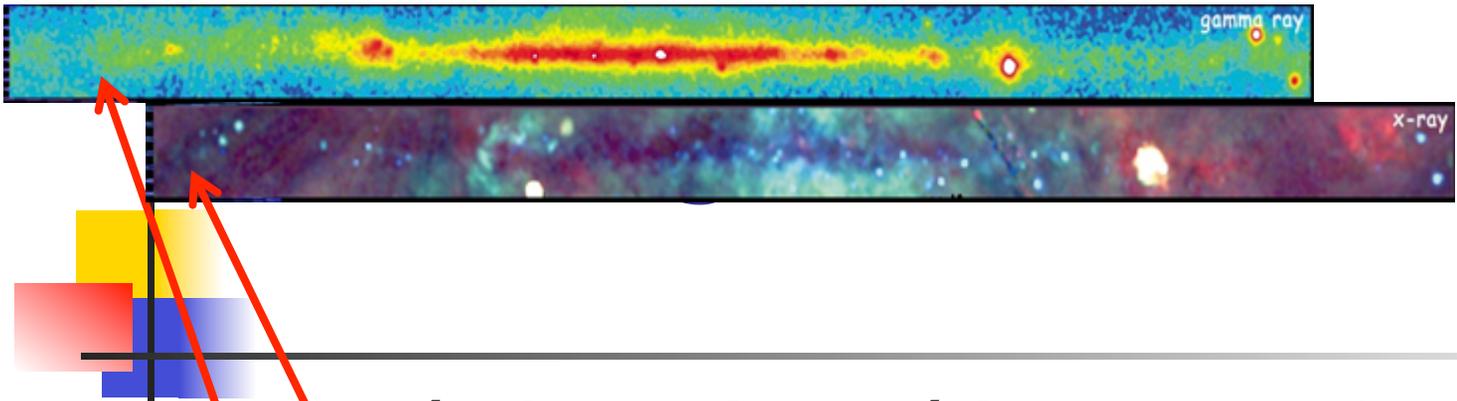
Astro-ingeniería

Astronomía observacional clásica - *espectro electromagnético*



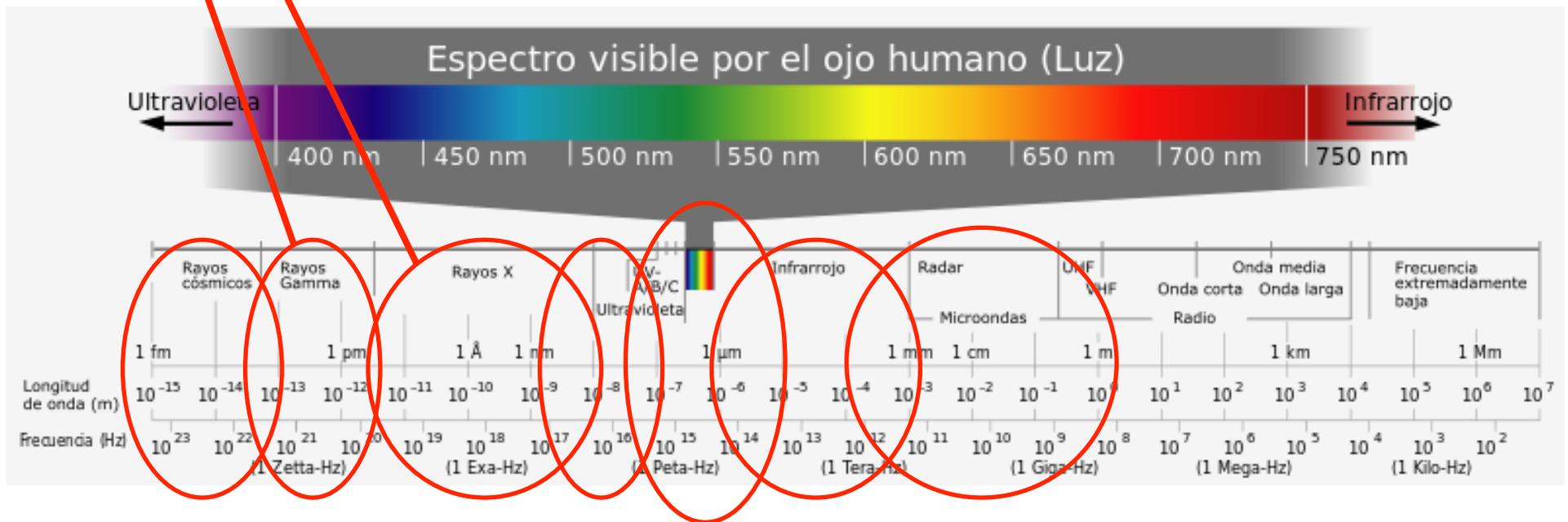
Instrumentos: extienden nuestra "vista" en las λ del espectro electromagnético que permite saber más sobre la estructura y dinámica del universo.

Astronomía de **múltiples** λ revela el universo en toda su luz.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astronomía observacional clásica - *espectro electromagnético*

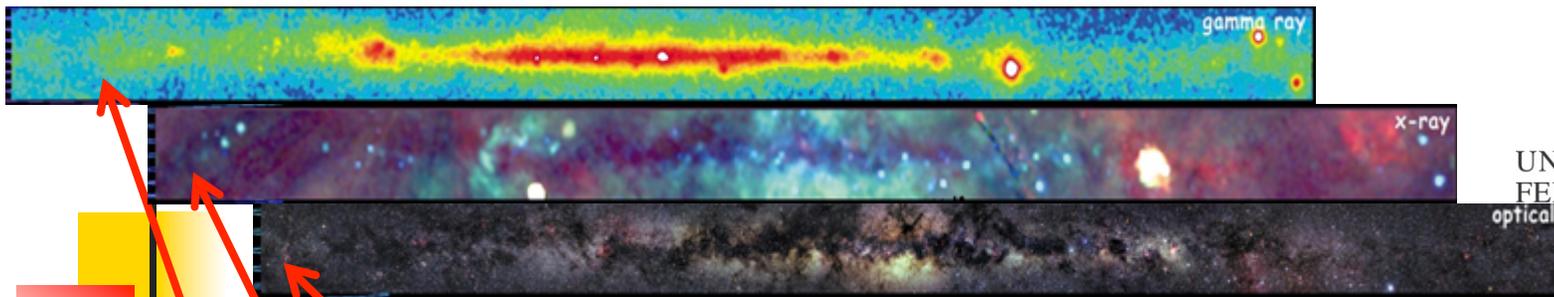


Instrumentos: extienden nuestra "vista" en las λ del espectro electromagnético que permite saber más sobre la estructura y dinámica del universo.

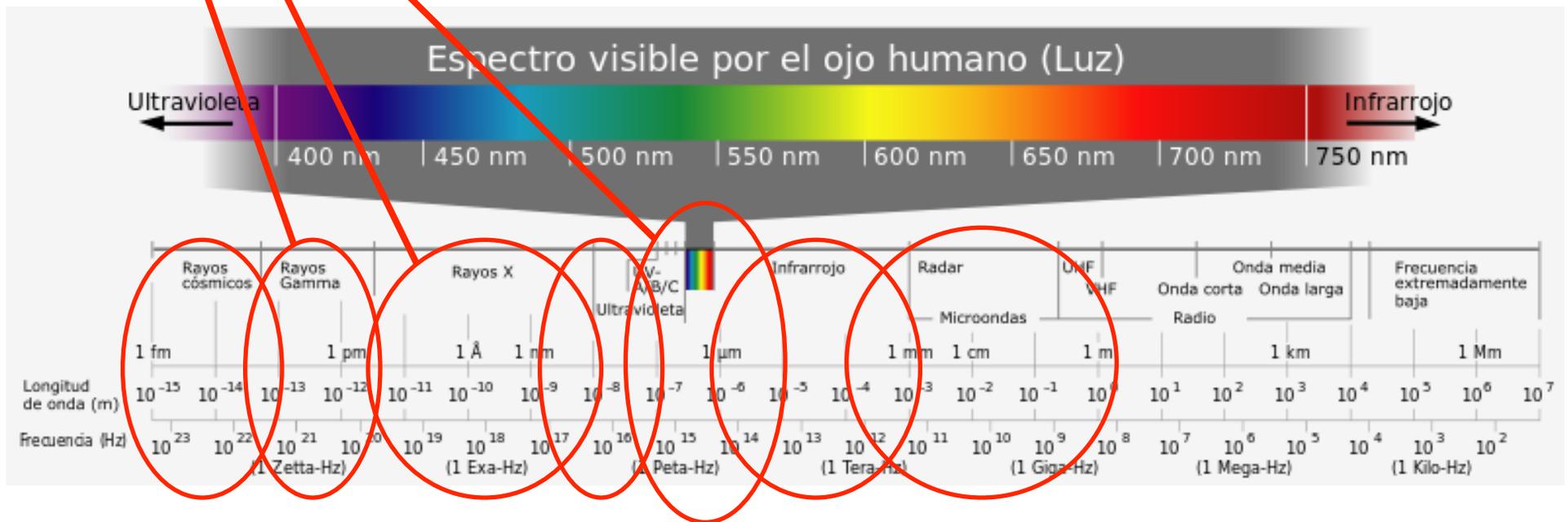
Astronomía de **múltiples** λ revela el universo en toda su luz.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA



Astronomía observacional clásica - *espectro electromagnético*

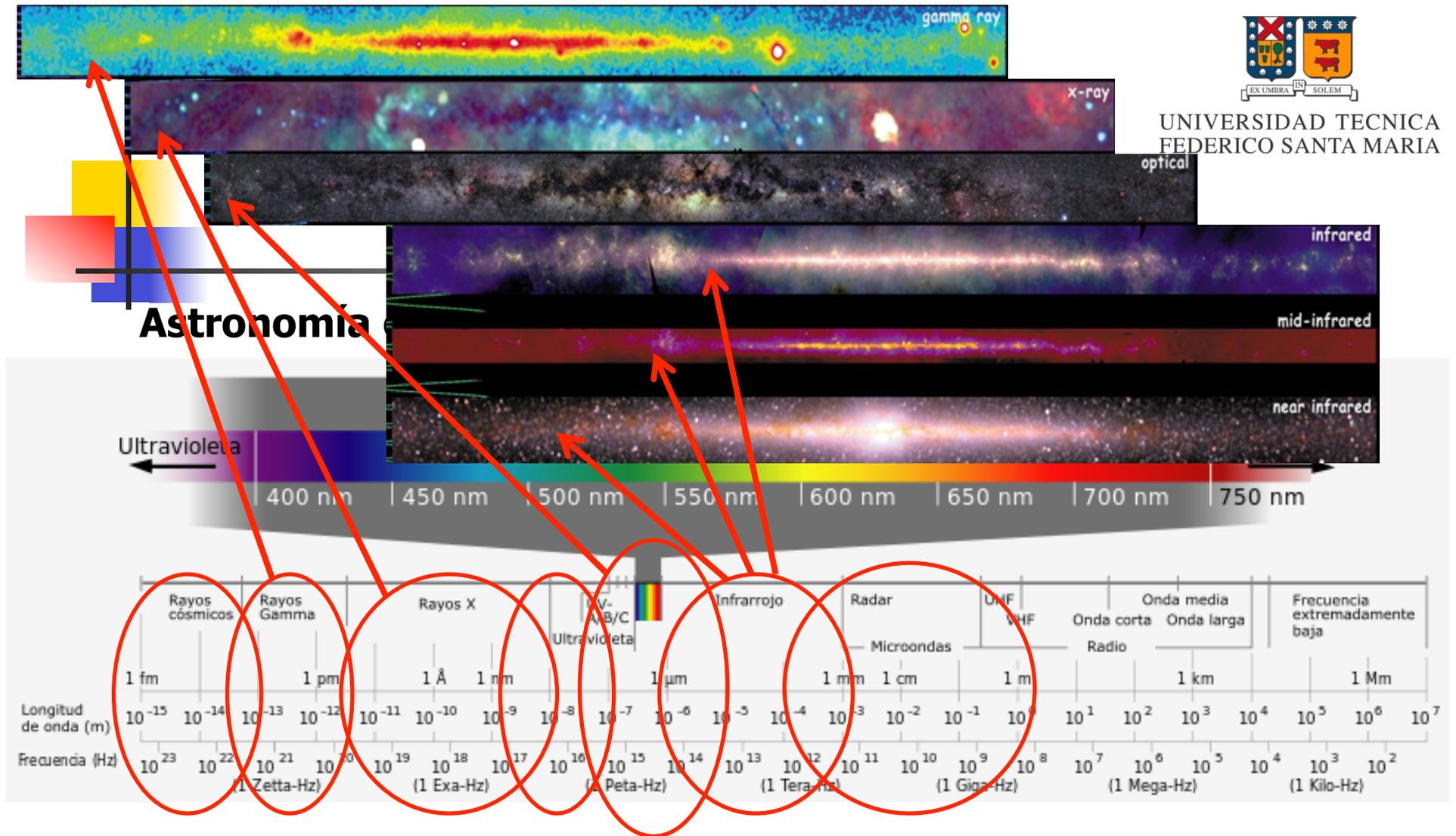


Instrumentos: extienden nuestra "vista" en las λ del espectro electromagnético que permite saber más sobre la estructura y dinámica del universo.

Astronomía de **múltiples** λ revela el universo en toda su luz.

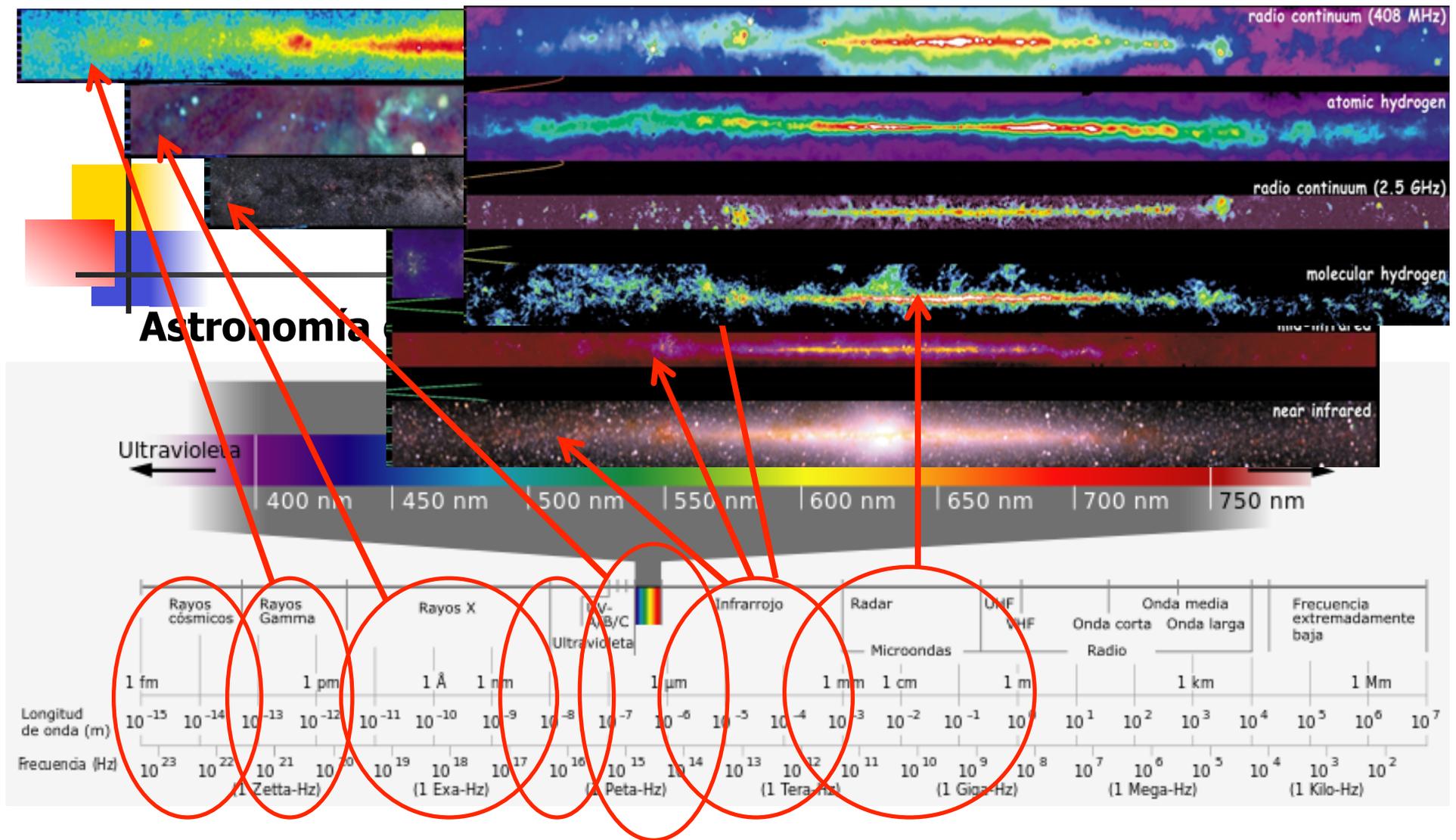


UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA



Instrumentos: extienden nuestra "vista" en las λ del espectro electromagnético que permite saber más sobre la estructura y dinámica del universo.

Astronomía de **múltiples** λ revela el universo en toda su luz.



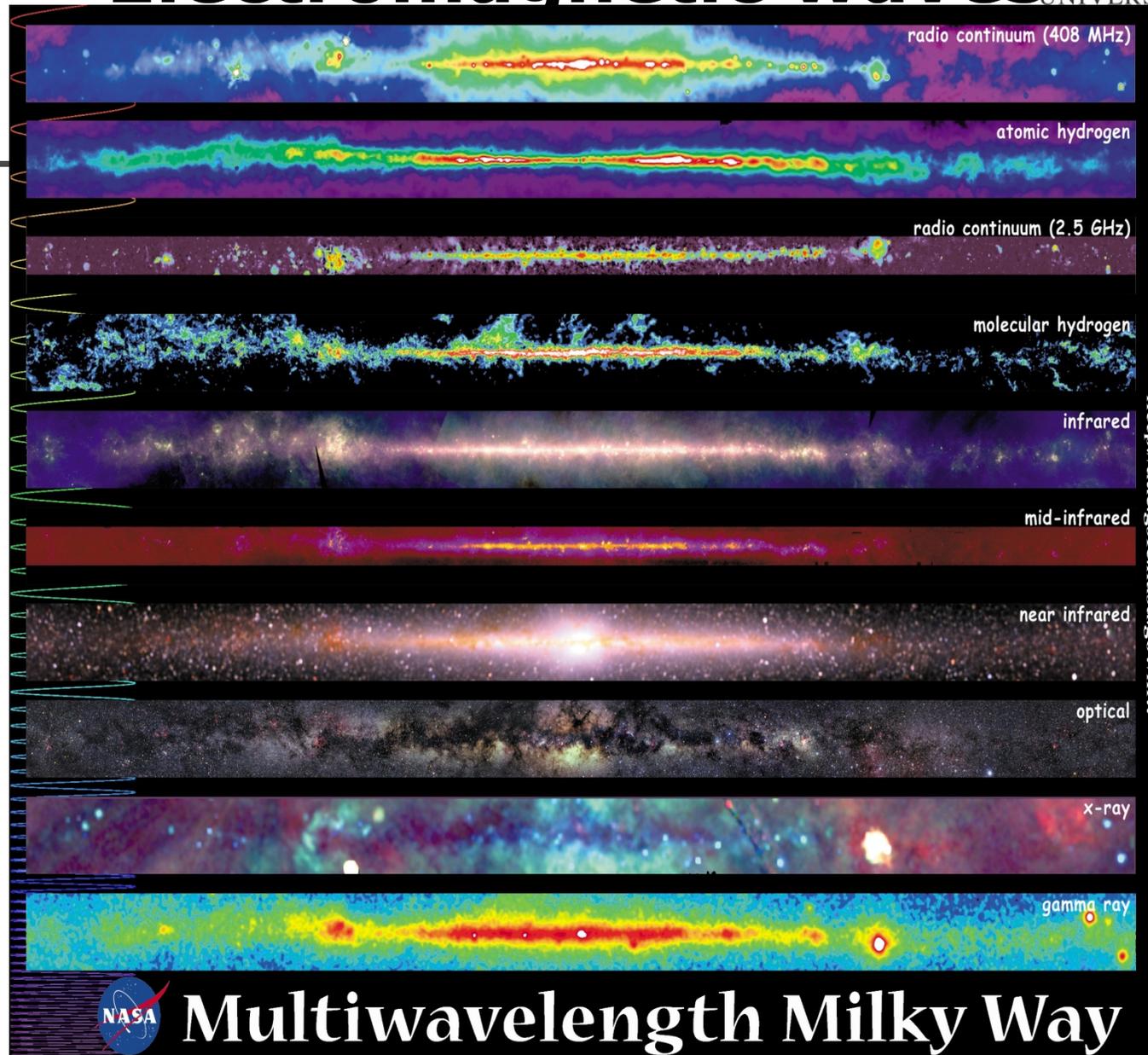
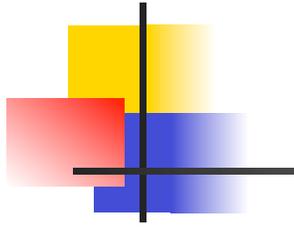
Instrumentos: extienden nuestra "vista" en las λ del espectro electromagnético que permite saber más sobre la estructura y dinámica del universo.

Astronomía de **múltiples** λ revela el universo en toda su luz.



UNIVERSIDAD TECNICA
SANTA MARIA

Electromagnetic waves



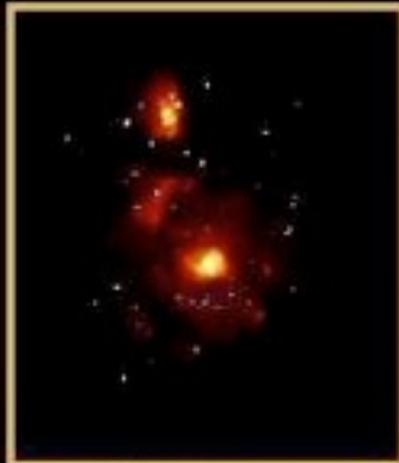


M51 – The Whirlpool Galaxy

Distance: 37,000,000 light-years (11 kpc)

Image Size = 11 x 13.5 arcmin

Visual Magnitude = 8.4



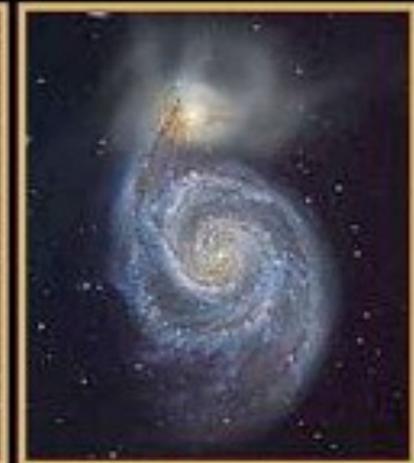
X-Ray: Chandra



Ultraviolet: GALEX



Visible: DSS



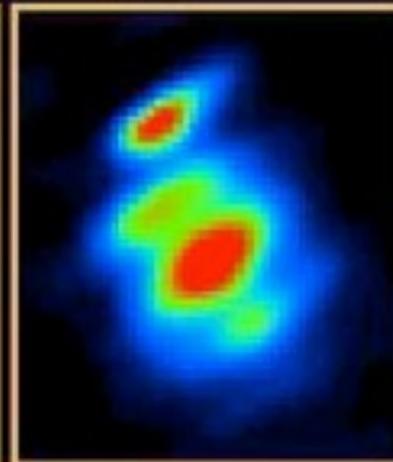
Visible: T. & D. Hallas



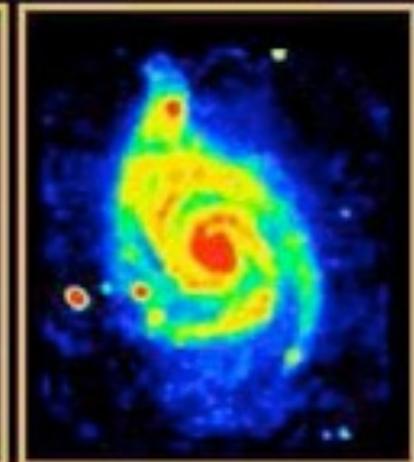
Near-Infrared: 2MASS



Mid-Infrared: ISO

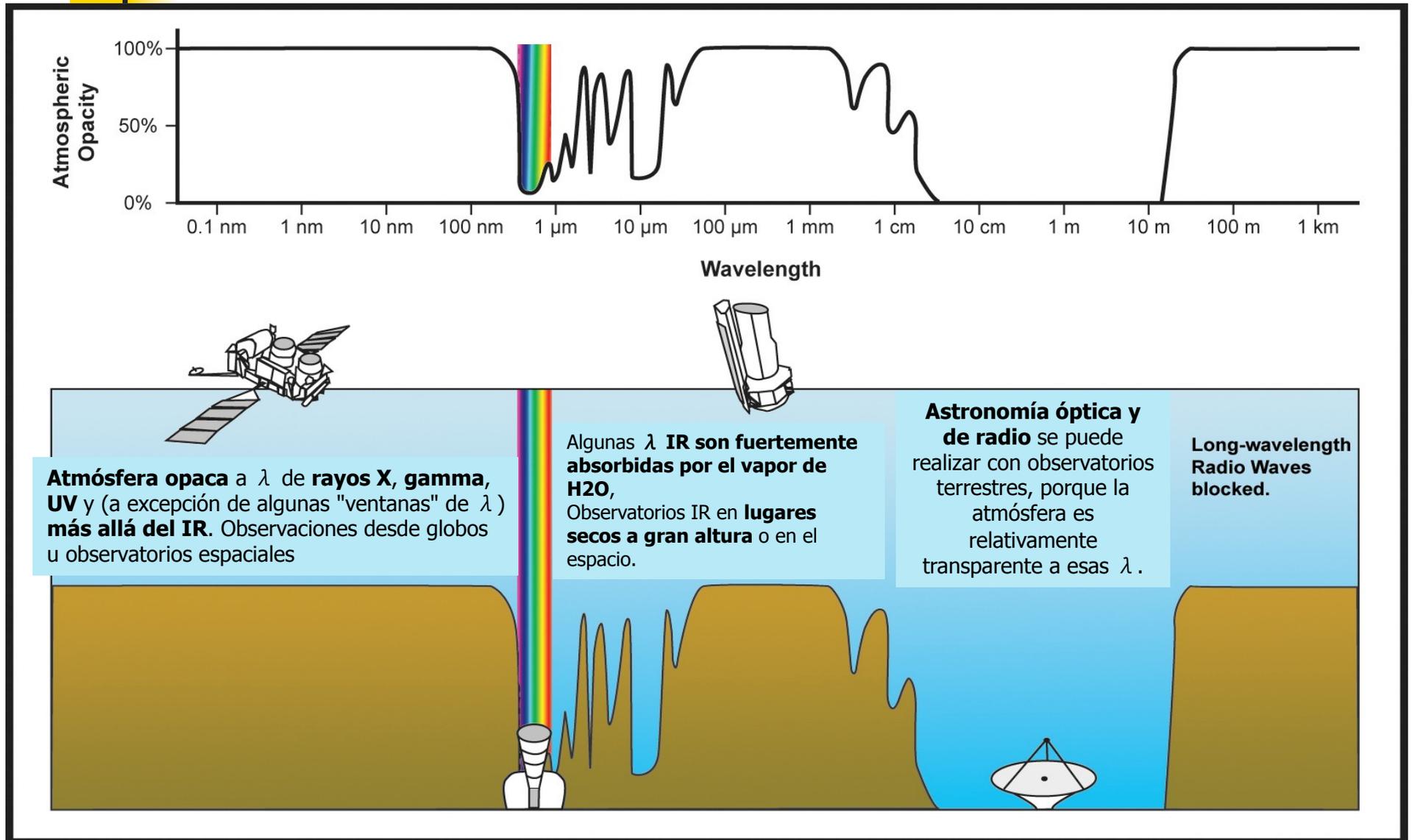


Far-Infrared: IRAS



Radio: VLA

Astro-ingeniería



Áreas de la Astronomía



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astrometría: Estudia posición de objetos en el cielo y su cambio de posición. Define sistema de coordenadas utilizado y cinemática de los objetos en nuestra galaxia.

Astrofísica: Estudio de la física del universo, incluyendo las propiedades de objetos astronómicos (luminosidad, densidad, temperatura, composición química).

Cosmología: Estudio del origen del universo y su evolución. El estudio de la cosmología es la máxima expresión de la astrofísica teórica.

Formación y evolución de galaxias: Estudio de la formación de galaxias y su evolución.

Astronomía galáctica: Estudio de estructura y componentes de nuestra galaxia y de otras.

Astronomía extragaláctica: Estudio de objetos fuera de la Vía Láctea.

Astronomía estelar: Estudio de las estrellas, su nacimiento, evolución y muerte.

Evolución estelar: Estudio de la evolución de las estrellas desde su formación hasta su muerte como un despojo estelar.

Formación estelar: Estudio de las condiciones y procesos que llevan a la formación de estrellas en el interior de nubes de gas.

Ciencias planetarias: Estudio de planetas del Sistema Solar y de planetas extrasolares.

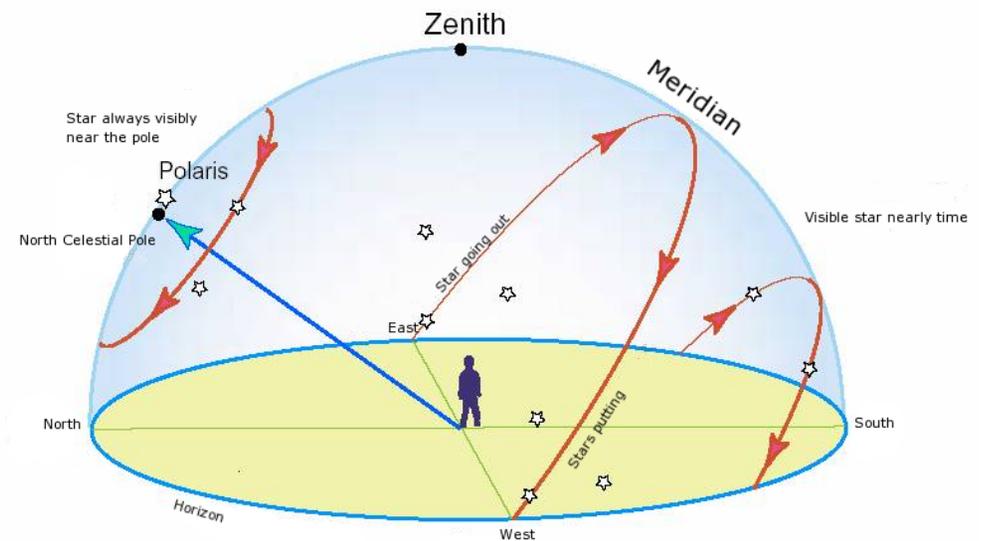
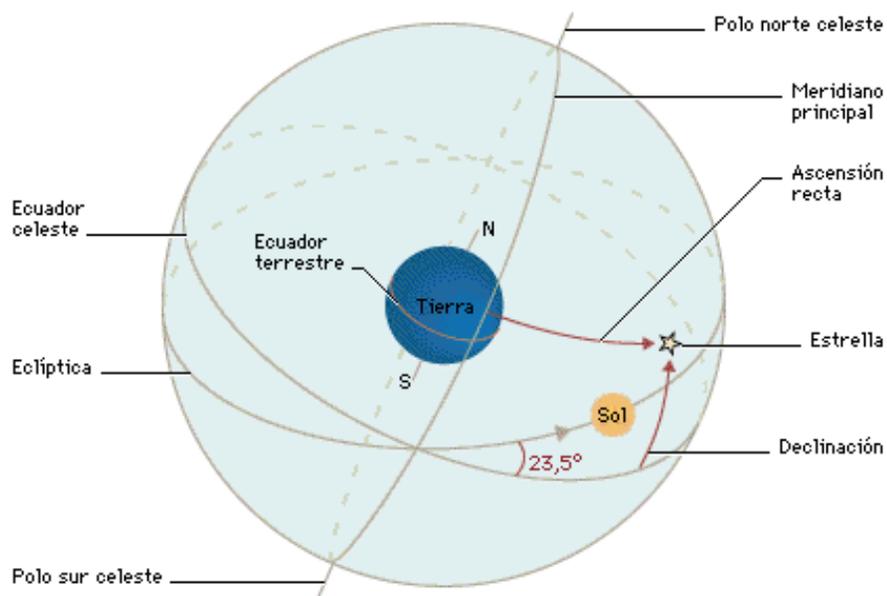
Astrobiología: Estudio de la aparición y evolución de sistemas biológicos en el universo.

Áreas de la Astronomía



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astrometría: Estudia posición de objetos en el cielo y su cambio de posición. Define sistema de coordenadas y cinemática de los objetos en nuestra galaxia.



Áreas de la Astronomía



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Astrometría: Estudia posición de objetos en el cielo y su cambio de posición. Define sistema de coordenadas utilizado y cinemática de los objetos en nuestra galaxia.

Astrofísica: Estudio de la física del universo, incluyendo las propiedades de objetos astronómicos (luminosidad, densidad, temperatura, composición química).

Cosmología: Estudio del origen del universo y su evolución. El estudio de la cosmología es la máxima expresión de la astrofísica teórica.

Formación y evolución de galaxias: Estudio de la formación de galaxias y su evolución.

Astronomía galáctica: Estudio de estructura y componentes de nuestra galaxia y de otras.

Astronomía extragaláctica: Estudio de objetos fuera de la Vía Láctea.

Astronomía estelar: Estudio de las estrellas, su nacimiento, evolución y muerte.

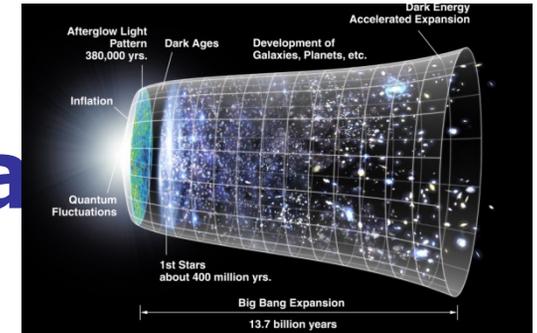
Evolución estelar: Estudio de la evolución de las estrellas desde su formación hasta su muerte como un despojo estelar.

Formación estelar: Estudio de las condiciones y procesos que llevan a la formación de estrellas en el interior de nubes de gas.

Ciencias planetarias: Estudio de planetas del Sistema Solar y de planetas extrasolares

Astrobiología: Estudio de la aparición y evolución de sistemas biológicos en el universo.

Áreas de la Astronomía

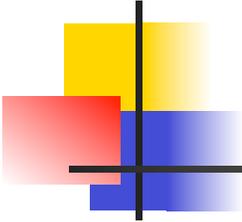


- Nacimiento y evolución de estructuras en el Universo (clusters de galaxias, evolución de agujeros negros):
 - ¿Cómo y cuándo se formaron las primeras galaxias?
 - ¿Cuál es la distribución de la materia oscura en los clusters de galaxias?
 - ¿Cómo se formaron los agujeros negros masivos en el centro de las galaxias?
- Poblaciones estelares en el Universo local (vía Láctea, etc):
 - ¿Cuál es la naturaleza y distribución de la materia oscura en las galaxias?
 - ¿Cómo se formó y evolucionó la Vía Láctea?
- La escala de distancias extragalácticas (Cefeidas, estrellas variables, etc):
 - ¿Cuáles son las mejores candelas para la determinación de distancias extragalácticas?
- Formación de estrellas (maternidades estelares, etc):
 - ¿Dónde/cómo se forman las estrellas? (<http://video.liberation.fr/video/cef1ed591d9s.html>)
- Planetas extrasolares y enanas marrones:
 - ¿Qué tipo de planetas se forman alrededor de otras estrellas?
- Supernovas y Energía Oscura:
 - ¿Cuál es la naturaleza de la energía oscura?



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Líneas Espectrales



Átomos: núcleo + electrones.

Núcleo: protones + neutrones unidos por **fuerza nuclear fuerte**

Protones: número de cargas positivas es la característica fisicoquímicas

Electrones: en órbitas permitidas (no en cualquier posición).

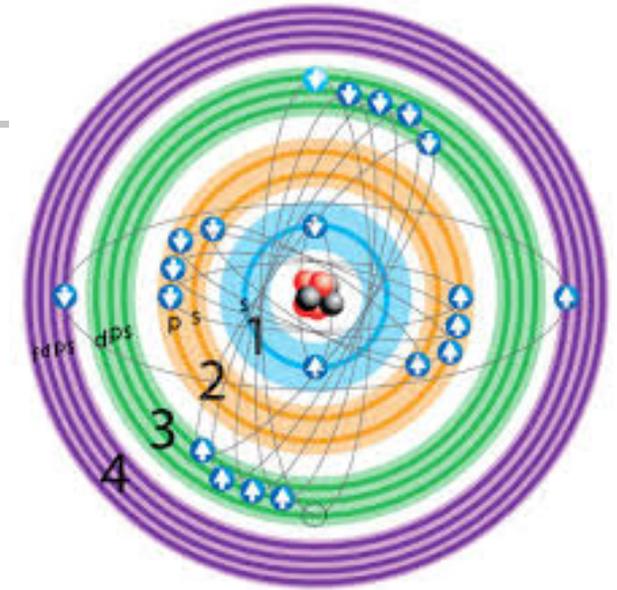
Electrón puede cambiar de una órbita a otra desocupada:

Al pasar a una más baja **emite energía**, que libera en forma de cuanto.

Al pasar a una más alta **absorbe energía** también en forma de cuanto de luz

Elemento químico: puede tener varios isótopos: número constante de protones pero no de neutrones

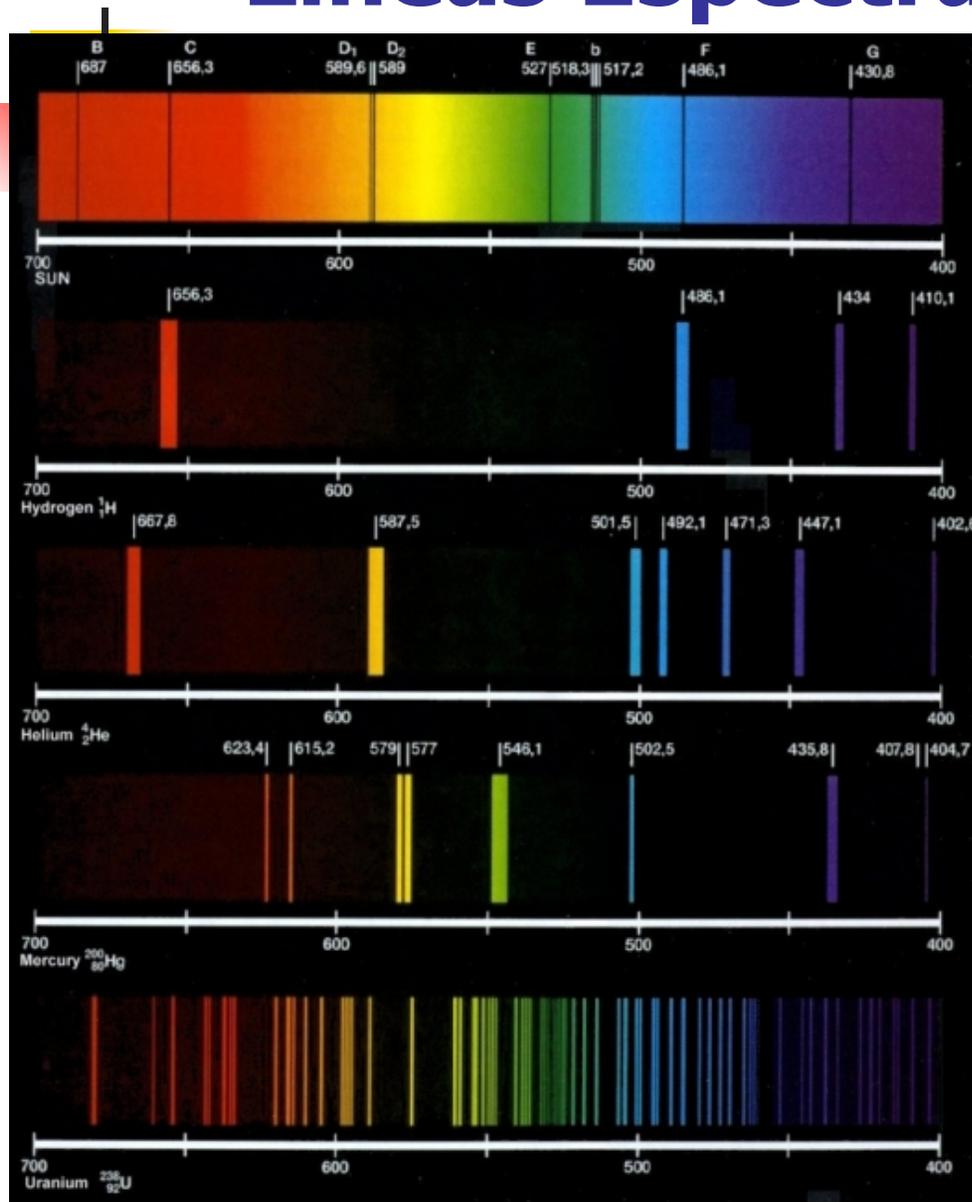
Isótopos del H: ^1H y ^2H (deuterio) y otro menos común es el tritio ^3H





UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Líneas Espectrales



Cuanto de luz emitido/absorbido:

especifico de cada órbita en c/átomo.

Energía electromagnética emitida/absorbida por un átomo determina tipo de átomo.

Material excitado (i.e. gas calentado por la luz estelar): multitud de átomos pueden estar sufriendo cambios en las órbitas de sus electrones, presentando gran cantidad de absorción y/o emisión de cuantos de energía. Estos fotones son "huellas" de identificación de átomos en el gas:

En el espectro de la luz de un gas o estrella se puede apreciar "**huecos**" en el espectro (líneas espectrales de absorción), que corresponden a las λ absorbidas por el átomo.

En el material incandescente se ve espectros con **líneas característicamente brillantes** a las que se denominan líneas de emisión.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Agenda

Introducción

- Astronomía en Chile

Astro-Ingeniería

- Áreas de la Astronomía

Técnicas Computacionales en Astronomía

- Data Mining, Machine Learning and HPC

Proyecto FONDEF de Astro-Informática

- IVOA, ChiVO, Team, Sponsors, etc.



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Técnicas computacionales en Astronomía

Áreas y sub-áreas:

Encontrar lo inusual:

Miles de millones de fuentes por noche:

Nova, supernova, GRBs:

Descubrimiento instantáneo:

Búsqueda de fuentes móviles:

Asteroides y cometas:

Movimiento propio de las estrellas:

Mapear la Vía Láctea:

Flujos de marea:

Estructura Galáctica:

La energía oscura y materia oscura:

Lentes gravitacionales:

Ligera distorsión en la forma:

Averiguar naturaleza de energía oscura:



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Técnicas computacionales en Astronomía

Áreas y sub-áreas: resuelta con

Encontrar lo inusual:

Miles de millones de fuentes por noche:
Nova, supernova, GRBs:
Descubrimiento instantáneo:

Búsqueda de fuentes móviles:

Asteroides y cometas:
Movimiento propio de las estrellas:

Mapear la Vía Láctea:

Flujos de marea:
Estructura Galáctica:

La energía oscura y materia oscura:

Lentes gravitacionales:
Ligera distorsión en la forma:
Averiguar naturaleza de energía oscura:

Técnicas computacionales:

Detección de anomalías
Reducción de dimensionalidad
Cruzamiento y matching de datos

Algoritmos de seguimiento (Tracking)
Filtros de Kalman

Estimación de densidad
Clustering (n-tuplas)

Visión por computador
Clasificadores débiles
Ajuste de Modelo Alta-D



Técnicas computacionales en Astronomía

- Funcionalidades para resolver problemas pueden catalogarse según:
 - Clasificación [**Predictivo**]
 - Clustering [*Descriptivo*]
 - Descubrimiento de regla de asociación [*Descriptivo*]
 - Descubrimiento de patrones secuenciales [*Descriptivo*]
 - Regresión [**Predictivo**]
 - Detección de Desviación [**Predictivo**]

- Funciones pueden implementarse con esquemas de aprendizaje **supervisado** o **no supervisado**, dando origen a modelos tales como:
 - MLP, GA, RBF, SVM, SVM-C, SOM, K-means, PPS, D-T, GM, ANN, etc...

Técnicas computacionales en Astronomía



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

- Separación de estrellas y galaxias:
 - Algoritmos como Redes Neuronales Artificiales (ANN), árboles de decisión (DT) y redes de mapas auto-organizados (SOM).
- Morfología de galaxias (distintas formas y tamaños):
 - **Secuencia Hubble**: Clasifica tipo de galaxias agrupándolas en categorías.
 - Algoritmo supervisado asigna una categoría a imagen de la cual se extraen parámetros.
 - Se puede aplicar ANN para predecir categoría de una galaxia con precisión similar al experto.
 - SOM y ANN: usadas para clasificar galaxias de imágenes del **HST**.
 - Enfoque alternativo: proyecto **Galaxy Zoo** empleó *crowdsourcing* para clasificación morfológica: se habilitó una aplicación para que el público general vea imágenes del *Sloan Digital Sky Survey* (SDSS) y las clasifique de acuerdo a un esquema trazado. En 6 meses más de 100 mil personas proporcionaron alrededor de 40 millones de clasificaciones para casi 900 mil galaxias.



Técnicas computacionales en Astronomía

■ Otras clasificaciones de galaxias:

- La población estelar de una galaxia es un factor determinante en sus propiedades físicas
- El espectro también puede utilizarse como método de clasificación.
- Se ha utilizado PCA (análisis de componente principal) para clasificación espectral.
- PCA también se usa en el pre-procesamiento de una clasificación espectral usando ANN.
- En el SDSS se usa ANN para predecir directamente los tipos espectrales.

■ Quásares y núcleos activos de galaxia (AGN):

- La mayoría de la radiación electromagnética en el Universo proviene de estrellas o de discos crecientes que rodean hoyos negros super-masivos en los AGN.
- AGN es importante en el caso de quásares, donde la luz de la región central puede eclipsar al resto de la galaxia.
- La selección de quásares y otros AGN desde BD astronómicas es un problema importante.
- DM:
 - Se ha usado ANN y DT para seleccionar quásares candidatos desde distintos estudios.
 - DT mejora fiabilidad de selección en un 85% comparada con un 60% obtenido con criterios simples.
 - Otros algoritmos: PCA, SVM, árboles KD, clustering, y estimación de densidad de kernel.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Agenda

Introducción

- Astronomía en Chile

Astro-Ingeniería

- Áreas de la Astronomía

Técnicas Computacionales en Astronomía

- Data Mining, Machine Learning and HPC

Proyecto FONDEF de Astro-Informática

- IVOA, ChiVO, Team, Sponsors, etc.

Equipo de Investigación



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

- **Director:** Mauricio Solar-USM (AI, HPC)
- **Alternate Director:** Marcelo Mendoza-USM (DB, DM)
- **Researchers Team**
 - Diego Mardones – UChile (Astronomer)
 - Guillermo Cabrera – UChile (DM, Astronomer)
 - Nelson Padilla – PUC (HPC, Astronomer)
 - Karim Pichara – PUC (DM)
 - Neil Nagar – UDEC (Astronomer)
 - Ricardo Contreras – UDEC (AI)
 - Victor Parada – USACH (AI)
- **Technology Transfer:** Gerardo Rivas.

Partners y Sponsors



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



ALMA

- Technology Transfer
- Sponsor (\$80.000.000 CLP):
 - Human Resources
 - Astronomers and Astro-Engineers
 - Infrastructure
 - Data generated by observing projects.



REUNA

- Communication infrastructure
- Sponsor (\$14.218.646 CLP):
 - Development
 - High Speed Connectivity



FONDEF

- Sponsor (US\$ 600 K)

Partners y Sponsors



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



UTFSM

- Principal (Astro-Informatic and Computer Science)
 - Software development



UNIVERSIDAD DE CHILE

U. de Chile

- Associate
- Responsibility (Astronomy and Computer Science)
 - Cluster and servers.



PUC de Chile

- Associate
- Responsibility (Astronomy and Computer Science)
 - Cluster dedicated to Astronomy.



U. de Concepción

- Associate
- Responsibility (Astronomy and Computer Science)
 - Applications in Astronomy.



Usach

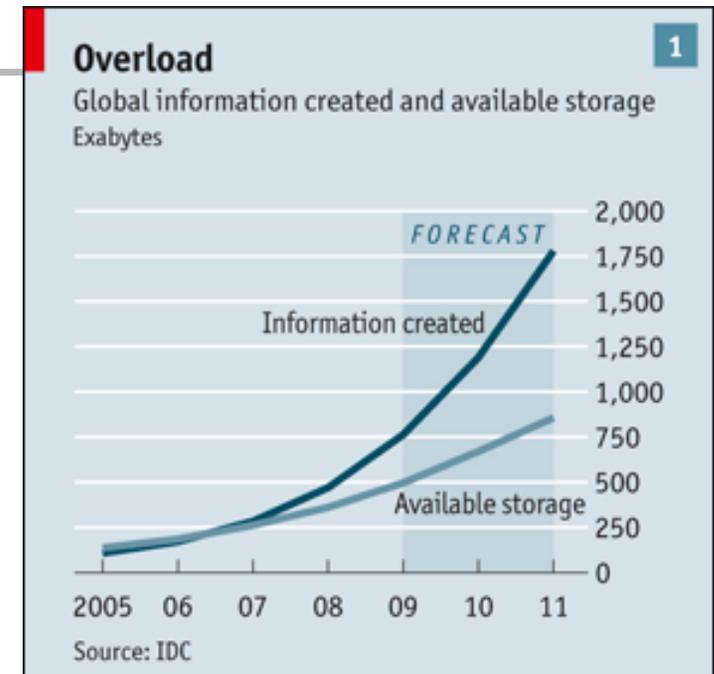
- Associate
 - Computer Science

Problem



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

- There is an increasing need to **store**, distribute and process the growing amount of **astronomical data** generated by observatories in Chile:
 - The large volume of data transmitted daily during life cycle of an observatory (≈ 30 years) to the storage and processing centers (data centers) of the respective observatories distributed mainly in the northern hemisphere (U.S., Europe and Japan).
 - The 10% of these data should be transmitted back (and potentially all data), demanding a high capacity data communication, which prevents proper scientific development of astronomers and researchers in the country, creating barriers to effective development of each investigation.
- There is no local capacity to manage and exploit large volumes of astronomical data



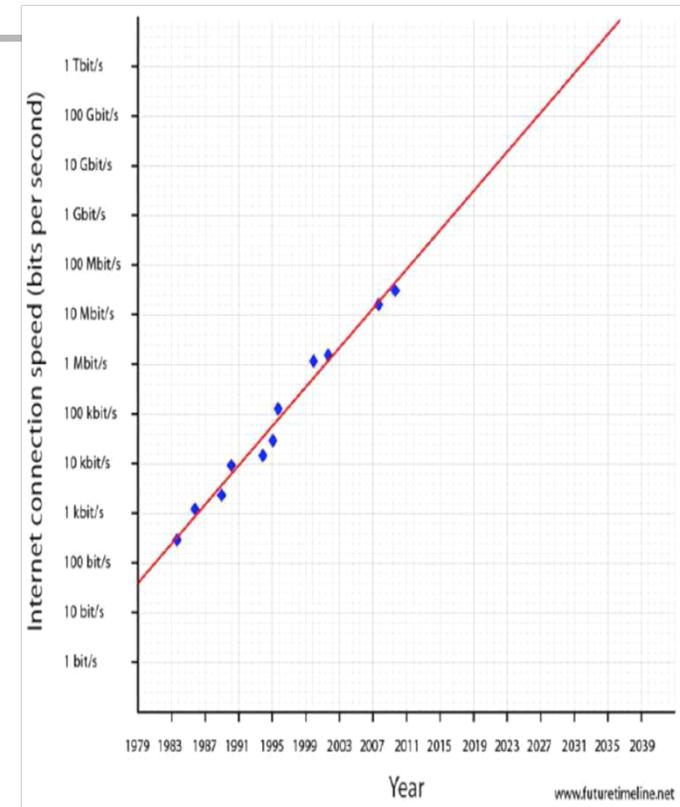
Problem



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

30 PB moves in 3.5 days at 1 Tb/s

- There is an increasing need to store, **distribute** and process the growing amount of **astronomical data** generated by observatories in Chile:
 - The large volume of data transmitted daily during life cycle of an observatory (≈ 30 years) to the storage and processing centers (data centers) of the respective observatories distributed mainly in the northern hemisphere (U.S., Europe and Japan) .
 - The 10% of these data should be transmitted back (and potentially all data), demanding a high capacity data communication, which prevents proper scientific development of astronomers and researchers in the country, creating barriers to effective development of each investigation.
- There is no local capacity to manage and exploit large volumes of astronomical data



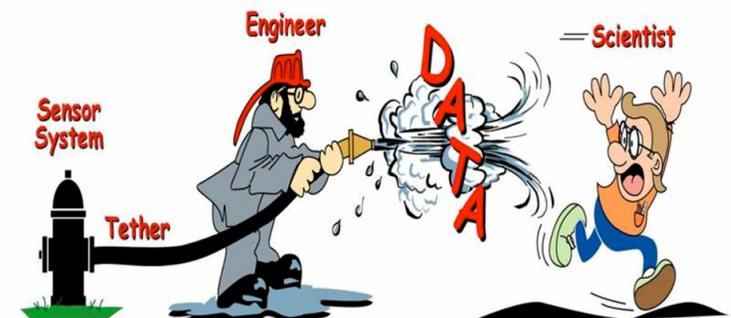
**Possible to store, but
too big to move**

Problem



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

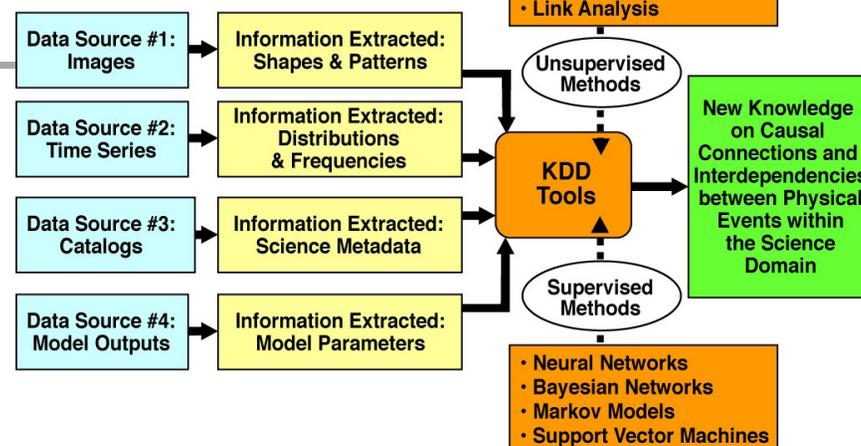
- There is an increasing need to store, distribute and **process** the growing amount of **astronomical data** generated by observatories in Chile:
 - The large volume of data transmitted daily during life cycle of an observatory (≈ 30 years) to the storage and processing centers (data centers) of the respective observatories distributed mainly in the northern hemisphere (U.S., Europe and Japan) .
 - The 10% of these data should be transmitted back (and potentially all data), demanding a high capacity data communication, which prevents proper scientific development of astronomers and researchers in the country, creating barriers to effective development of each investigation.
- There is no local capacity to manage and exploit large volumes of astronomical data



Solution



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

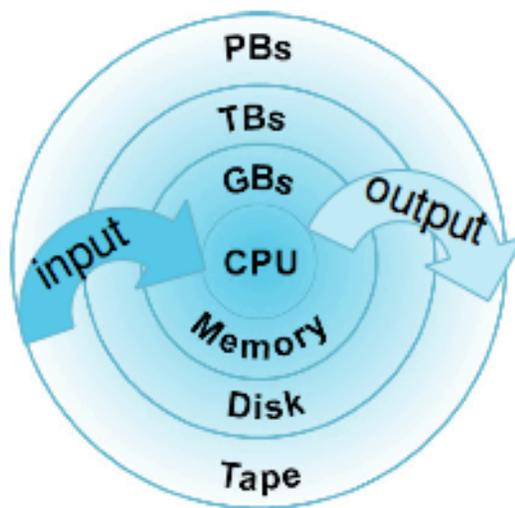


- Research, analyze, design and implement a platform for **distributed** Virtual Observatory (VO), to locally support Chilean astronomers in the collection and storage of data. This platform also includes the development of **efficient algorithms** for semantic search and intelligent processing of large-scale data from the VO.

- Differential attributes of our proposal:
 - This is a local development with leading universities in different areas (DB, AI, HPC, DM, Astronomy), which integrates different capabilities (NLHPC, Astro-PUC Center, UTFSM) for a solution for astronomical observatories (ALMA).
- Mechanisms and working methods previously used in successful projects in astronomy applications (ESO, NRAO, AUI, ALMA, etc).
- How will we address the problem:
 - Establish **partnerships** between data centers, observatories and research to create a VO.
 - **Implement distributed platform** for storing large-scale data with high performance information retrieval and semantic search.
 - Develop new data mining algorithms and computational intelligence information extraction (manual and automatic).

Big Data Systems Require a *Data-centric Architecture* for Performance

Old Compute-centric Model



Data lives on disk and tape
Move data to CPU as needed
Deep Storage Hierarchy

Manycore



FPGA



Massive Parallelism
Persistent Memory

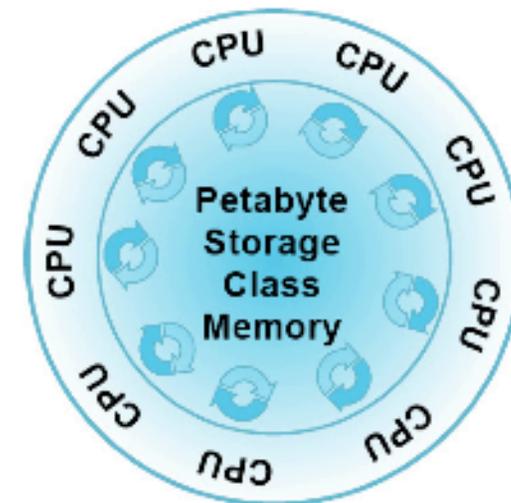


Flash



Phase Change

New Data-centric Model



Data lives in persistent memory
Many CPU's surround and use
Shallow/Flat Storage Hierarchy

Largest change in system architecture since the System 360
Huge impact on hardware, systems software, and application design

Solution: hypotheses



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

- To develop a **VO**, semantic search algorithms on the data (labeled by astronomers), and integration of data mining techniques (DM) and Artificial Intelligence (AI) with HPC, which allows **large-scale data processing** in an efficient and effectively way.

a) Scientific Hypothesis:

- Detecting **data with relevant information** to the search process performed on distributed astronomical data, excluding trivial data, that is realizable through **prediction methods, classification and segmentation**, normally used in the field of AI as well as in DM that considers centralized data and medium-scale.

b) Technological (and cost) Hypothesis:

- Create a **virtual observatory** based on distributed data, generates technology to data storage and manages significant amounts of information derived from astronomical observations. Also enables experts to process data with low computational time, wherever they are, as long as they have internet connection.

c) Other Hypothesis (work style between universities and partners):

- The implementation of a VO of astronomical data distributed in several universities, **collaborative work**, allows local technological development of each center and also increases the Technology transfer from the centers to the **users** of such information.

Usuarios y Beneficiarios



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

- The **VO** is a great **benefit** for all astronomers in Chile, as they can access astronomical DBs which support its research, with historical and recent data. The **VO** will be free, so everybody can freely access.
May 2015 - Proposal for (or Agreed to) Italy (Options are Venice/Trieste/Sesto)
Fall 2015 – Proposal for Australia, also Armenia offered (Areg Mickaelian via ei)
- Among the beneficiaries is the **Chilean community** in general, as schools get closer to astronomy in Chile, placing greater emphasis on rural schools, to contribute to the **learning** of this science.
- **Goal of this project:** The aim of this project is related to the design and implementation of a VO, which shall comply the standard of the **International Virtual Observatory Alliance** (IVOA), in closely collaboration with the Chilean astronomical community.
- **Chilean Virtual Observatory (ChiVO):** member of IVOA (ivoa.net)

IVOA: Int. VO Alliance



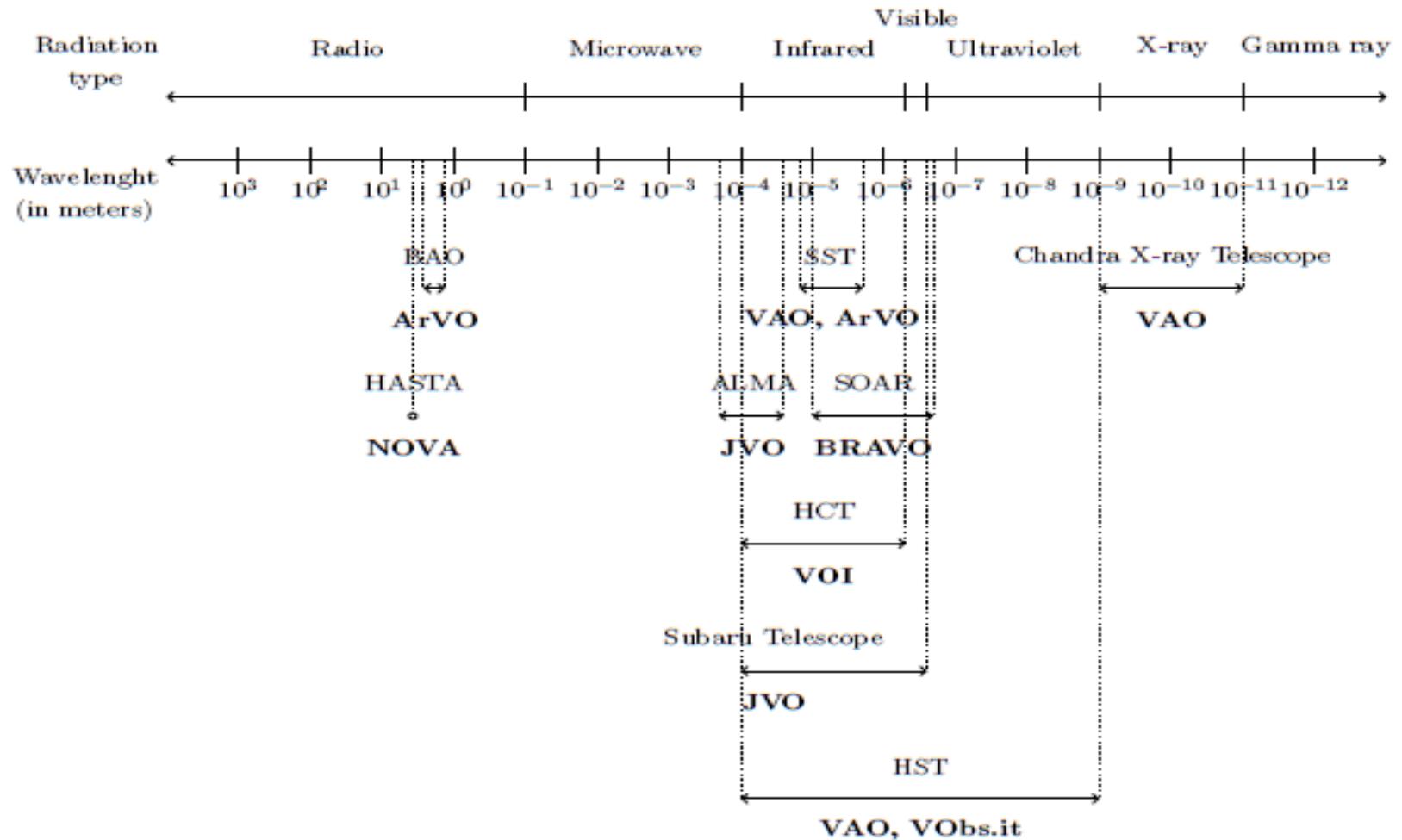
UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Project	URL
Argentina VO	http://nova.conicet.gov.ar/
Armenian VO	http://www.aras.am/Arvo/arvo.htm
AstroGrid	http://www.astrogrid.org/
Australian VO	http://aus-vo.org.au/
Brazilian VO	http://www.lna.br/bravo/
Canadian VO	http://www.cadc-ccda.hia-ihp
Chinese VO	http://www.china-vo.org/
European Space Agency	http://www.sciops.esa.int/index.php?
European VO	http://www.euro-vo.org/
German Astrophysical VO	http://www.g-vo.org/
Hungarian VO	http://hvo.elte.hu/en/
Italian VO	http://vobs.astro.it/
Japanese VO	http://jvo.nao.ac.jp/
OV France	http://www.france-vo.org/
Russian VO	http://www.inasan.rssi.ru/eng/rvo/
Spanish VO	http://svo.cab.inta-csic.es/
Ukranian VO	http://www.ukr-vo.org/
Virtual Astronomical Observatory,	http://www.usvao.org/
VO India	http://vo.iucaa.ernet.in/~voi/

IVOA: Int. VO Alliance



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



IVOA: Level 0



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Capa de usuarios:

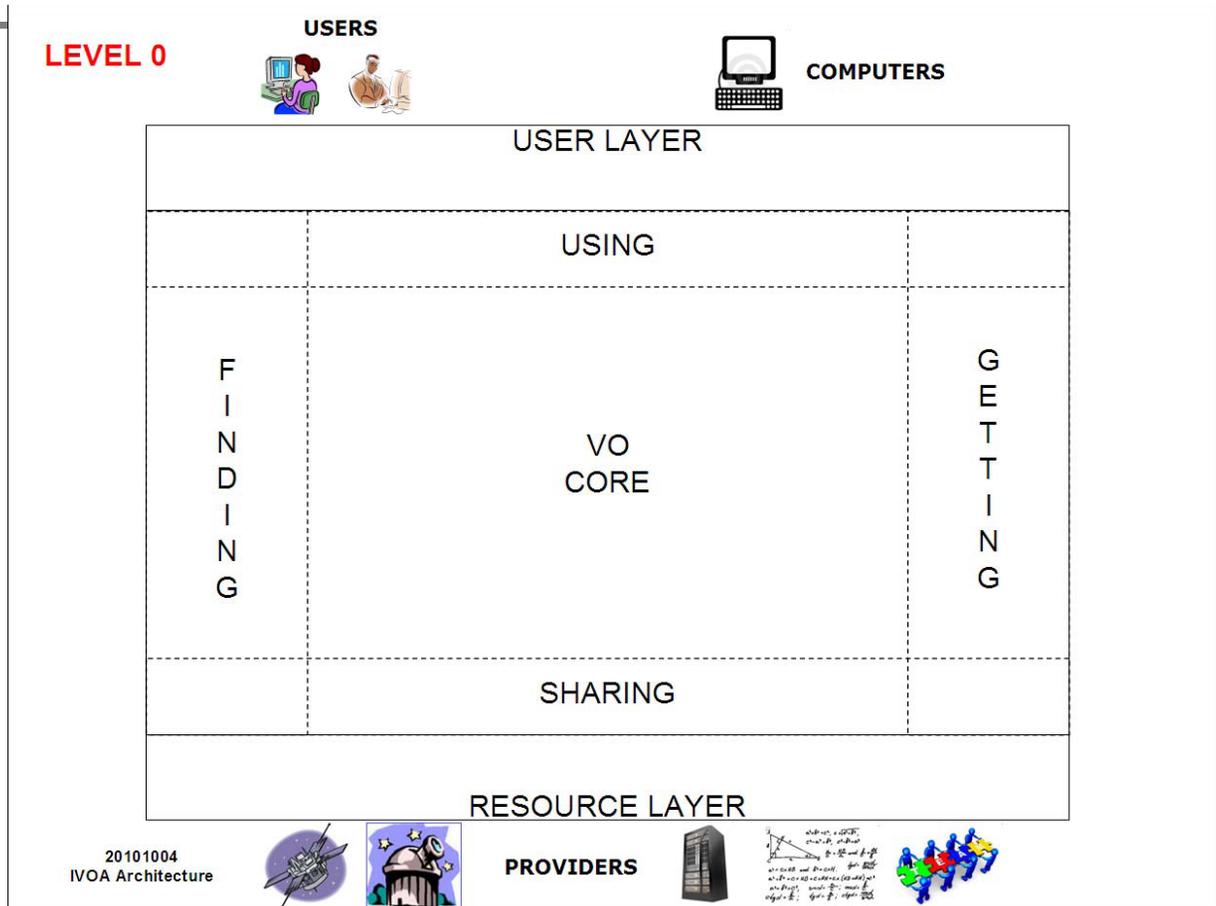
investigadores que buscan consumir datos.

Capa intermedia:

es la capa que permite conectar las dos capas anteriores de manera transparente para los investigadores. Esta interacción se puede llevar a cabo buscando/obteniendo datos.

Capa de recursos:

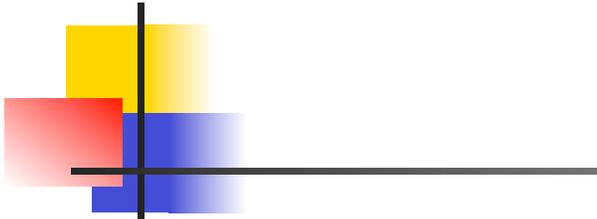
compilado de datos astronómicos provenientes de distintos instrumentos.



IVOA: Level 1



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



Capa de usuarios:

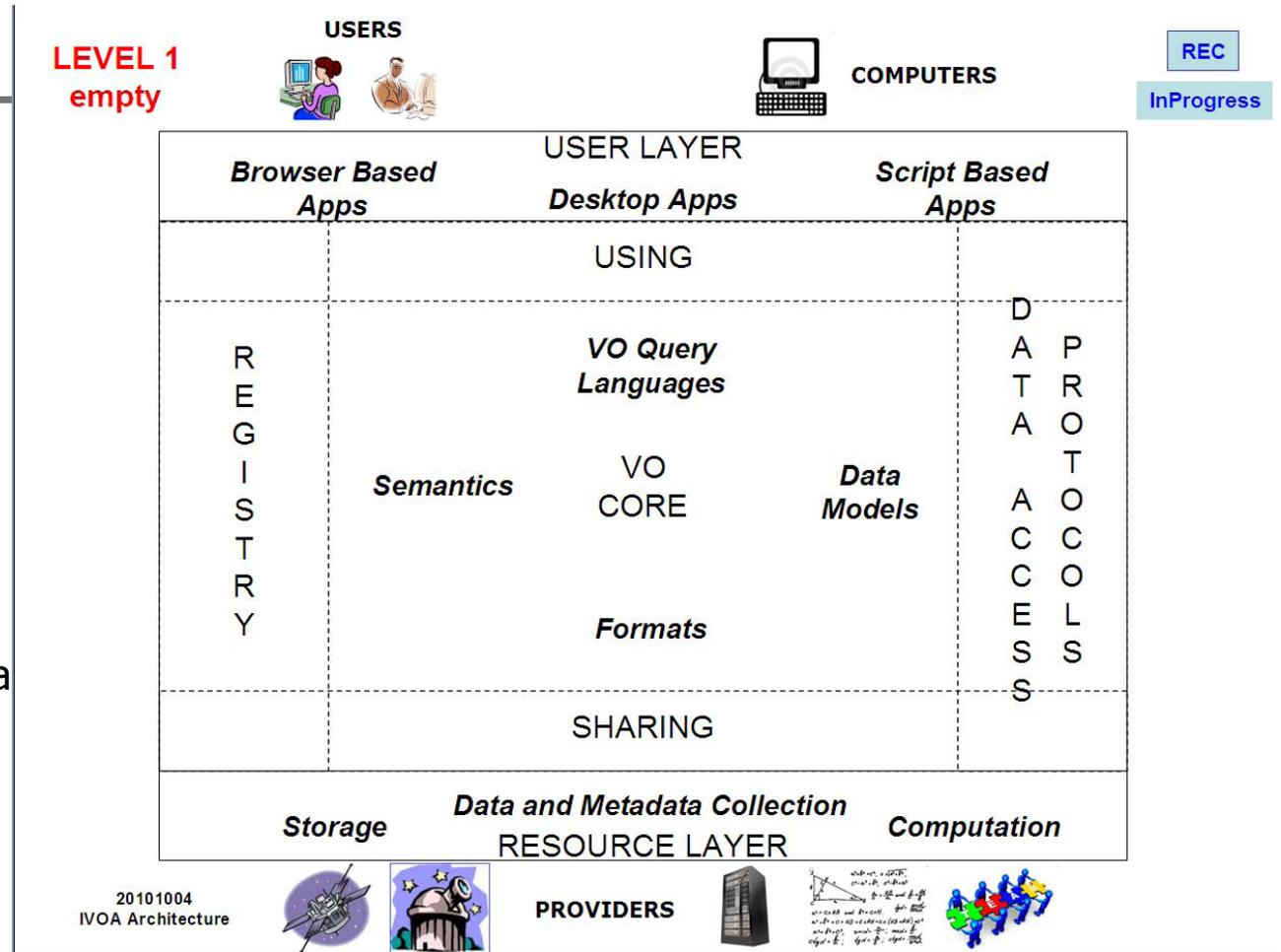
Un consumidor puede querer acceder a los datos desde un navegador, escritorio, o mediante un script.

Capa intermedia:

crea un framework para compartir los datos, compuesto por VOQL, Data Models, Semantics, Formats.

Capa de recursos:

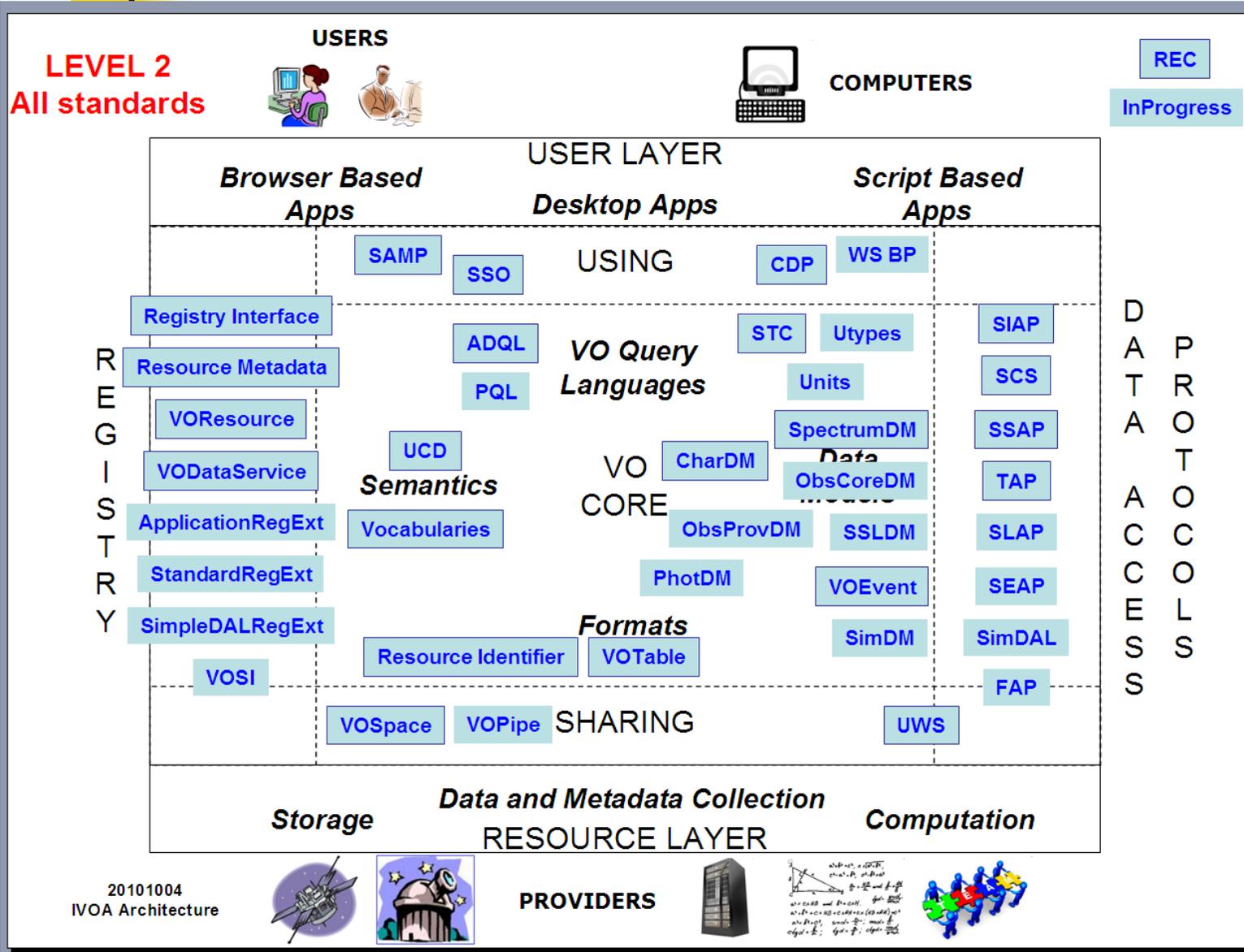
está compuesto de colección de datos y provenientes de distintos servidores.



IVOA: Level 2



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



La arquitectura nivel 2 es el VO regido por estándares y protocolos de IVOA.

Cada estándar se relaciona con la capa a la que pertenece.

IVOA: Level 2



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

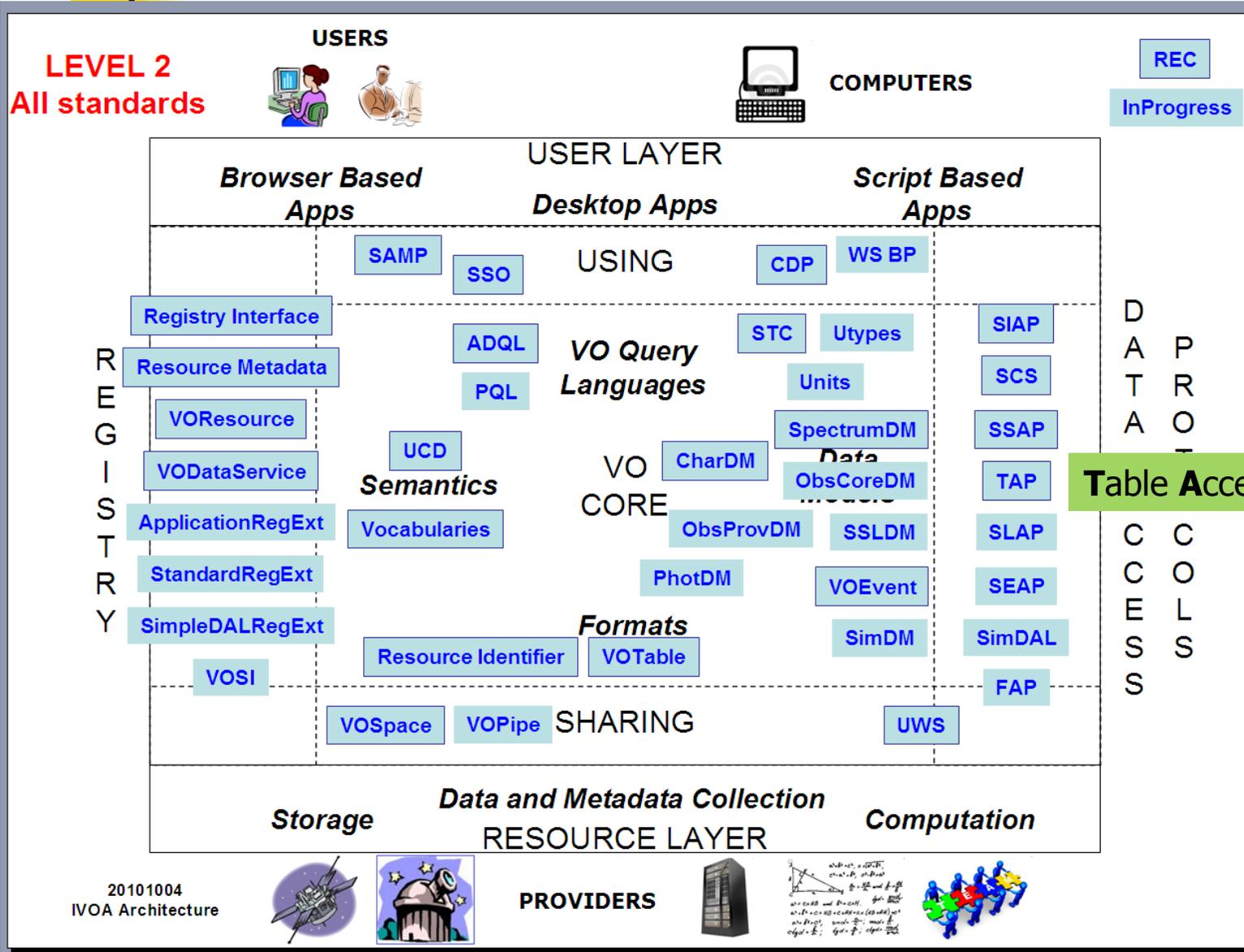
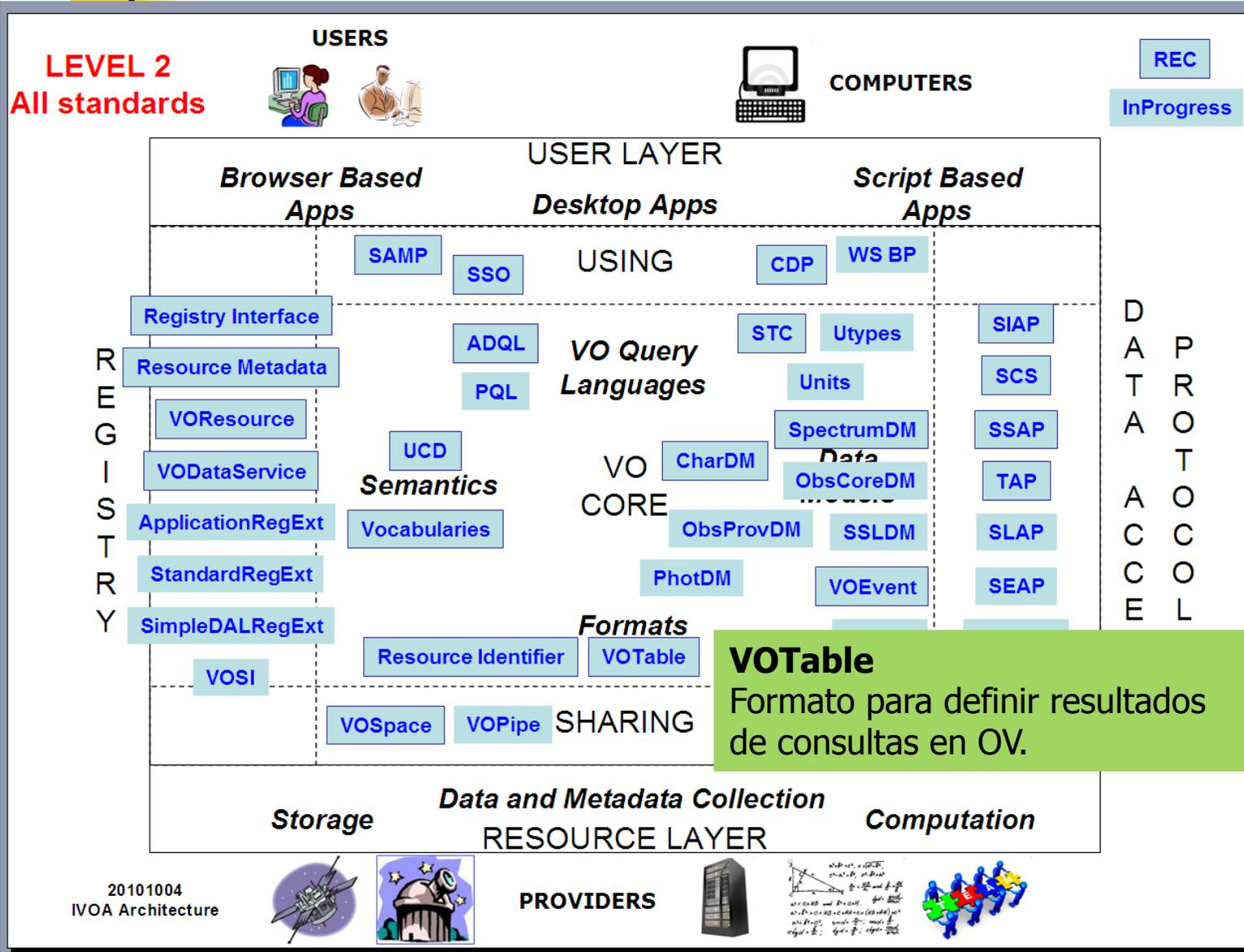


Table Access Protocol

IVOA: Level 2



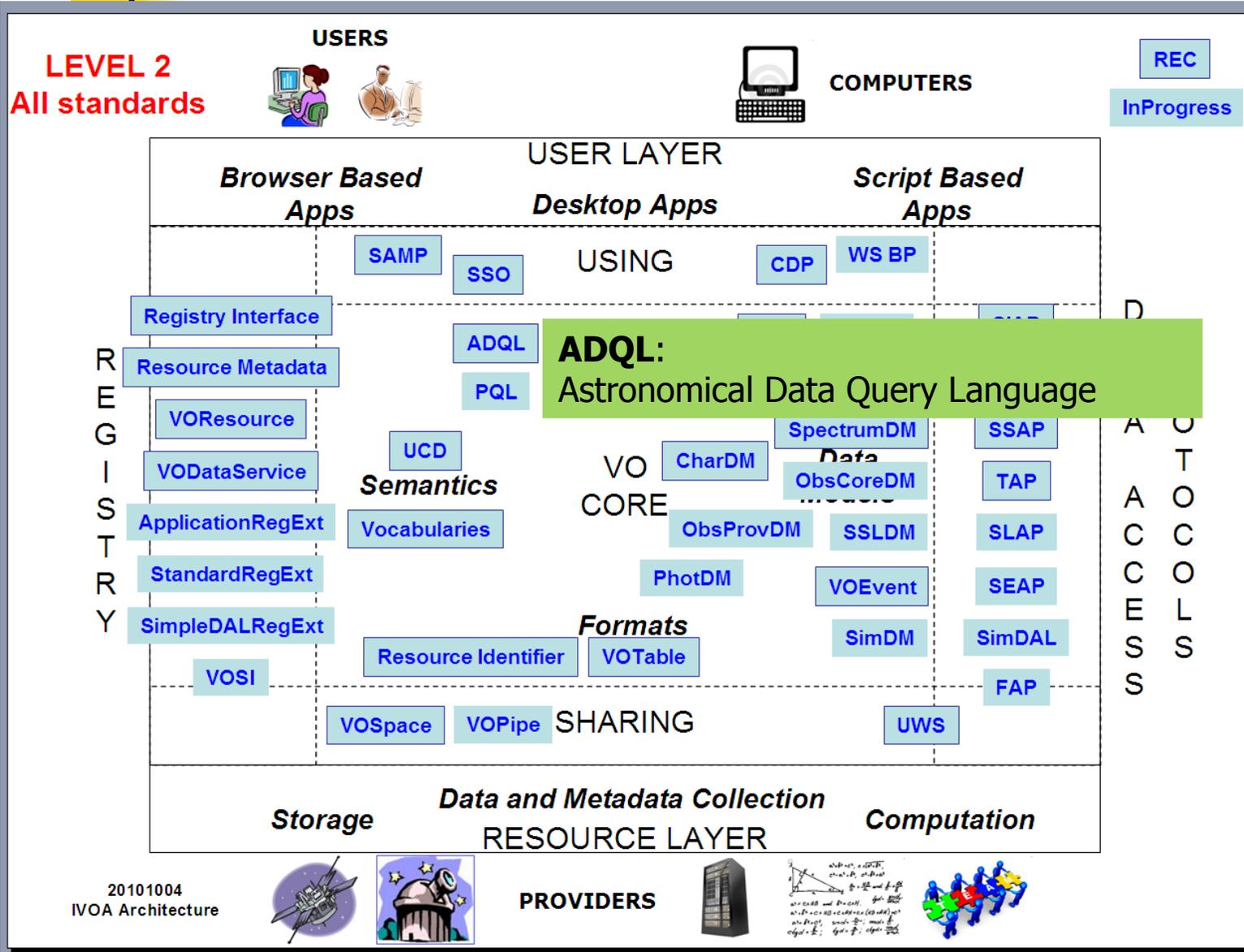
UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



IVOA: Level 2



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



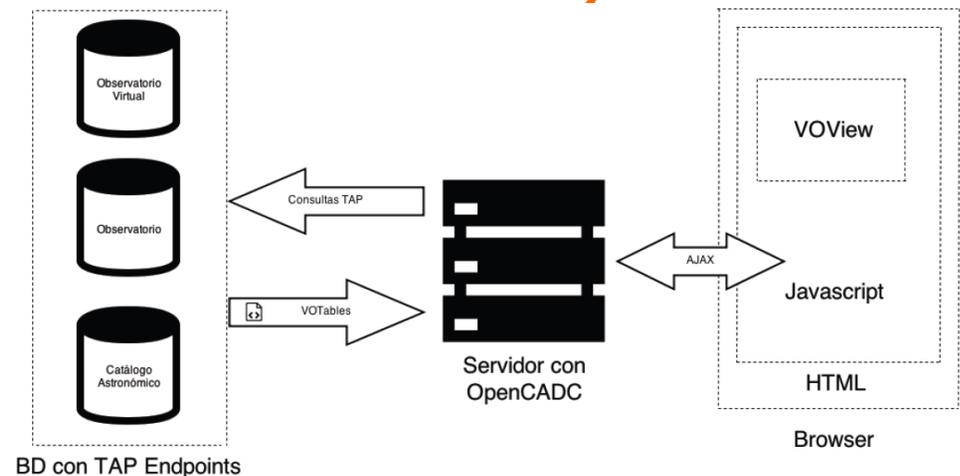
ChiVO requirements



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Search by:

1. **Coordinates or sky region**
2. **Object name or type**
3. **Spectral metadata (frequency and resolution)**
4. **Spatial metadata (angular resolution and field of view)**
5. **Temporal metadata**
6. **Polarization**
7. **Cross search**
8. **Simulations**
9. **Bibliographic resources**



Requirement classification:

Need: Essential, desirable, optional.

Temporal priority: High, medium, low.

Chilean VO = ChIVO



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Astro Form VO

http://www.astroformvo.com

VO Query Form

Form | ADQL | Upload your FITS

Object	Observation	Time
<input type="text" value="Object name"/>	<input type="text" value="Observation ID"/>	<input type="text" value="Observation date"/>
<input type="text" value="Ra Dec"/>	<input type="text" value="P.I. Name"/>	<input type="text" value="Integration time"/>
<input type="text" value="Search radius"/>	<input type="text" value="Proposal Keywords"/>	<input type="text" value="Time span"/>

Problems we are solving



- Using the stacking technique with ALMA data**
 - USACH, Prof. V. Parada, Student: Rodrigo Jara
- Automatic classifier for objects in ALMA**
 - PUC de Chile, Prof. K. Pichara, Student: Andrés Riveros
- An automatized astronomical structure detection and classification tool for images outside the visible range**
 - U. de Chile, Prof. E. Vera , Student: Giselle Font
- Classifying galaxies, detecting mergers of galaxies and other morphologies**
 - Univ. de Concepción, Prof. R. Contreras, Student: Luis Gabriel Salazar
- Index generation of astronomical objects using images**
 - UTFSM, Prof. M. Mendoza, Student: Gabriel Candia
- Conceptual design of a VO**
 - UTFSM, Prof. M. Solar, Student: Walter Fariña
- Image Mining**
 - UTFSM, Prof. M. Solar, Student: Cristian Fuentes
- Search for stellar objects in DB**
 - UTFSM, Prof. M. Solar, Student: Rodrigo Gregorio



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Gracias!!

Whatever else astronomy may or may not be
who can doubt it to be the most beautiful of
the sciences?

Isaac Asimov

