

# 大數據星系天文學

黃崇源

中央大學天文所

# 大數據

---

✓ 何謂大數據？

✓ 一般大數據的定義：

– 3V：容量(Volume)、快速(Velocity)、與  
多樣性(Variety)

– 4V：真實性(Veracity)

✓ 個別特性與總體性質的表現！

✓ 統計性質

✓ 相關性

# 多大的資料量才算大數據？

- ✓ 數據的量不是重點！
- ✓ 使用大數據的目的是要從資料中找到無法從小數據找到的結果！
- ✓ 所需資料的大小與研究的內容有關
- ✓ 大數據的重點：多樣性和覆蓋率！
- ✓ 大數據的結果常讓我們看到一些無法從小數據預期的發現！

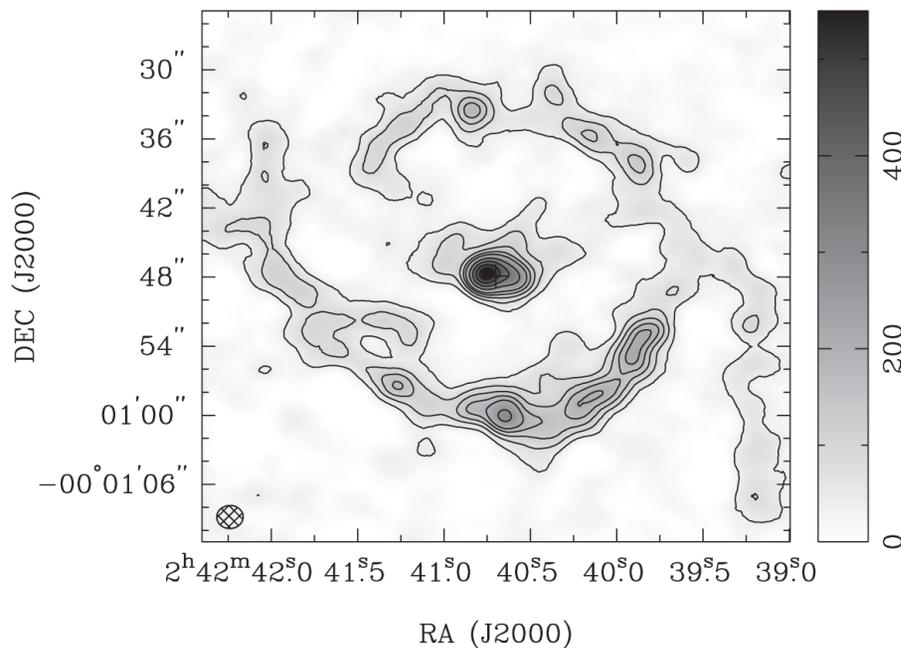
# 傳統天文學(相對於大數據天文學)

---

- ✓ 提高望遠鏡的靈敏度
  - 看得愈暗
  - 看得愈遠
- ✓ 提高望遠鏡的解析度
  - 看得更仔細
- ✓ 使用不同波段的望遠鏡
- ✓ 發現新的現象找到新的結果！
- ✓ 有時傳統天文學的數據量也是很大！

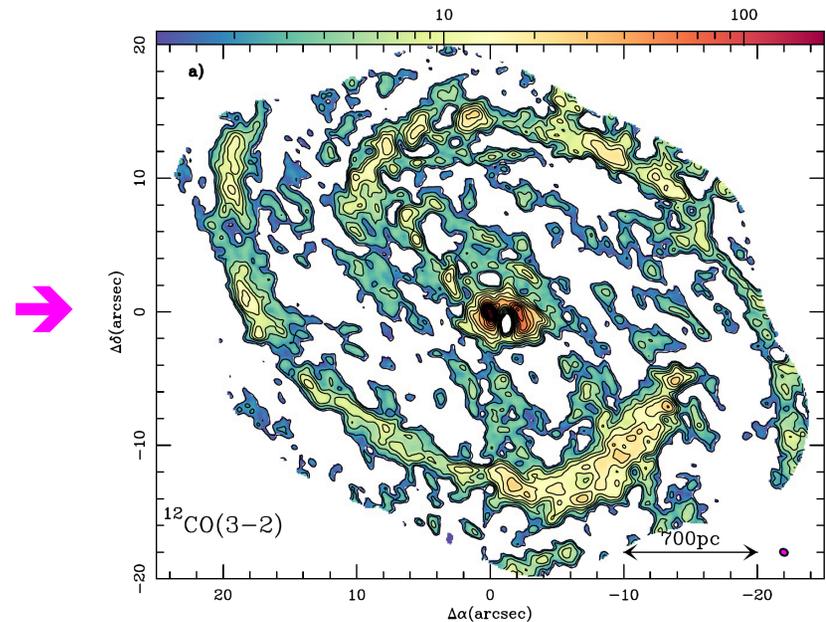
# 例子: NGC 1068 的 CO(3-2) 觀測

SMA



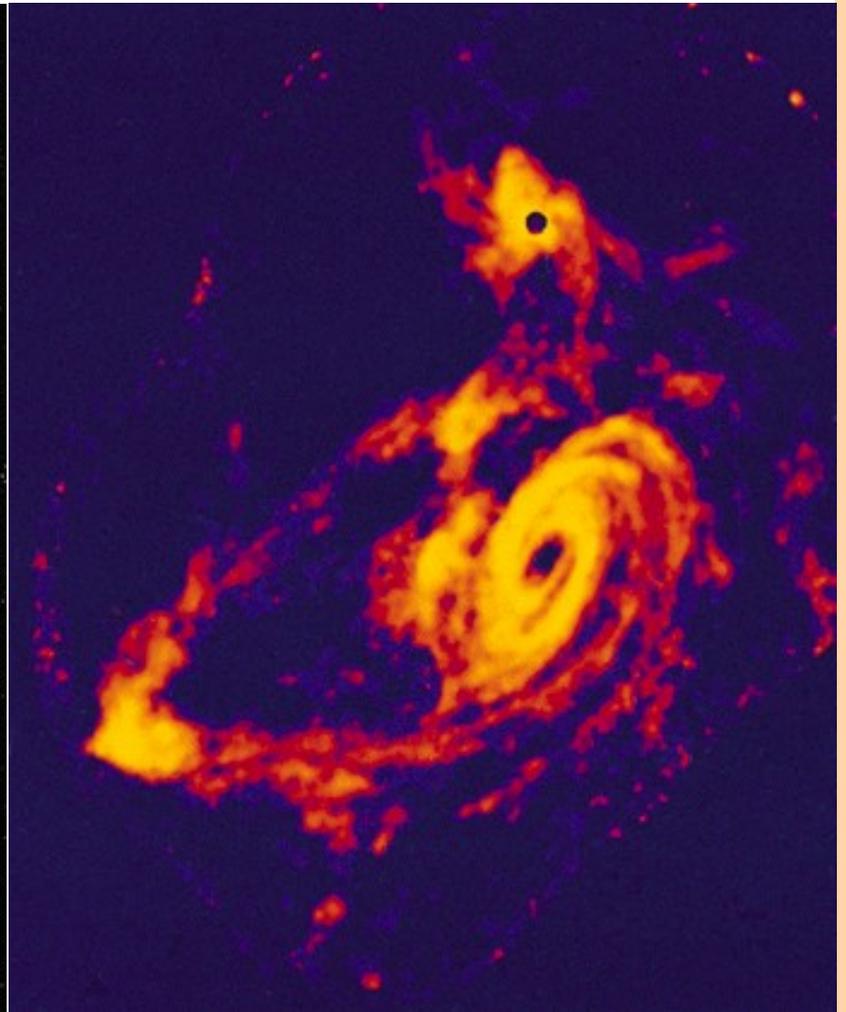
(*Tsai et al 2012*)

ALMA



(*Garcia-Burillo et al 2014*)

# 例子: M81的可見及21公分譜線

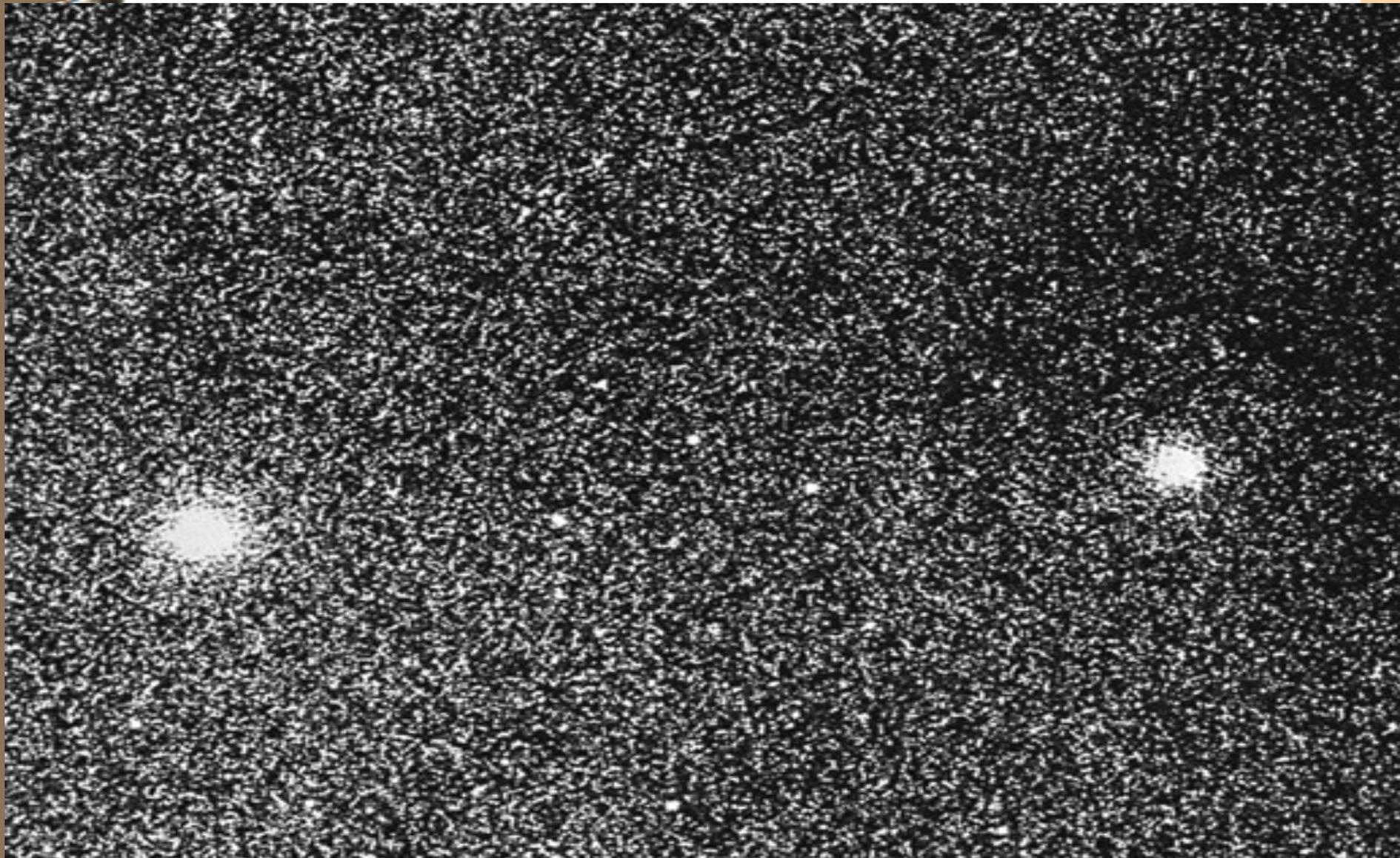


# 大數據星系天文？

---

- ✓ 星星與星系
- ✓ 如果有一台聚光能力很強，且可以有  
很長的曝光時間，可以看到更暗更遠  
的天體，我們會看到什麼？
- ✓ 滿天星星還是滿天星系？
- ✓ 看你所看的天空方向！

# 滿天星星？



# 滿天星系！

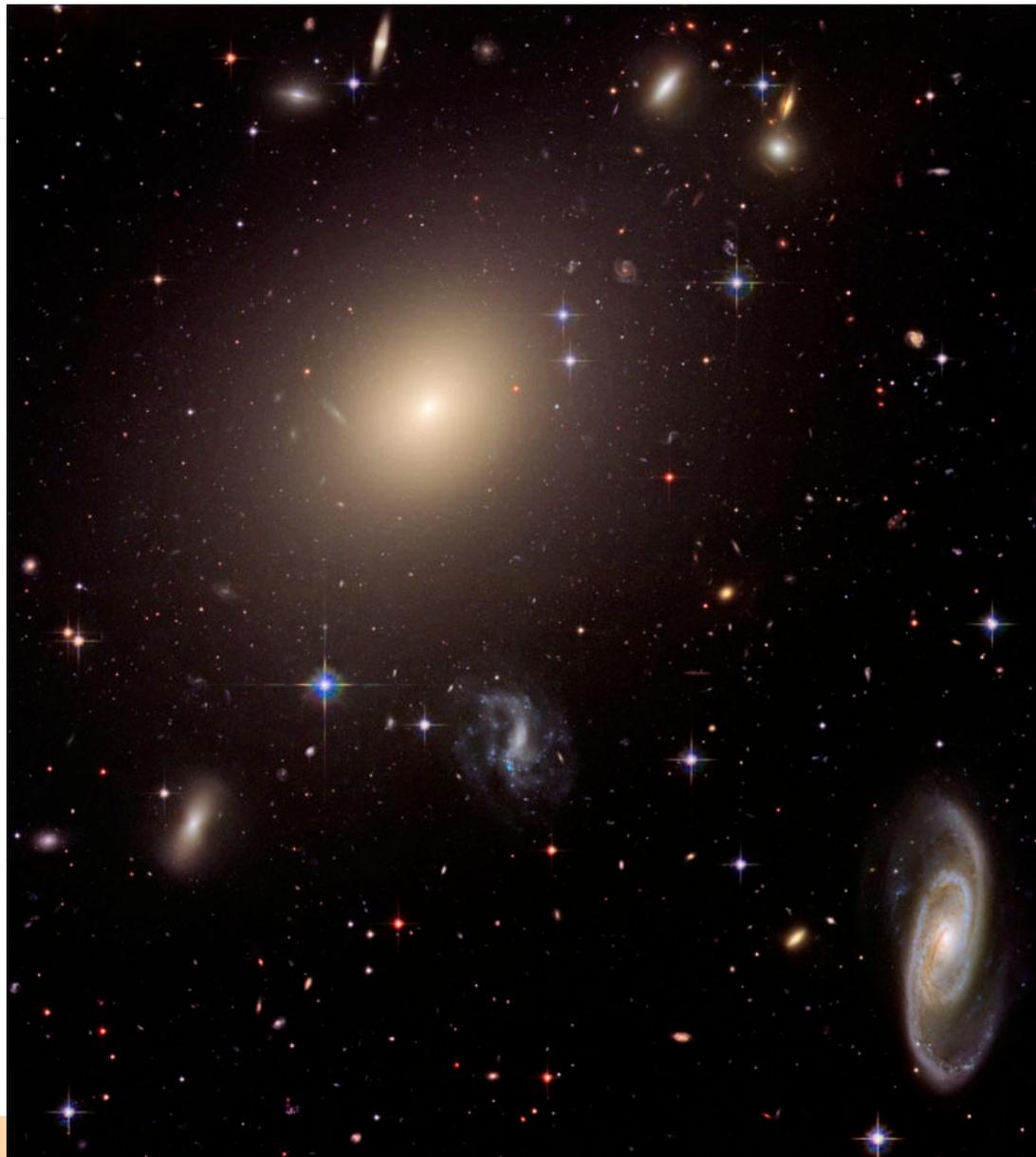


# 星系形態的多樣性

---

- ✓ 我們的銀河約有一千億顆星星
- ✓ 可觀測的宇宙中也約有一千億個星系
- ✓ 星星的形態：圓形
- ✓ 星系的形態：多樣
- ✓ 大數據的星系可幫助我們研究星系形態

# 星系形態的多樣性



# 各式星系

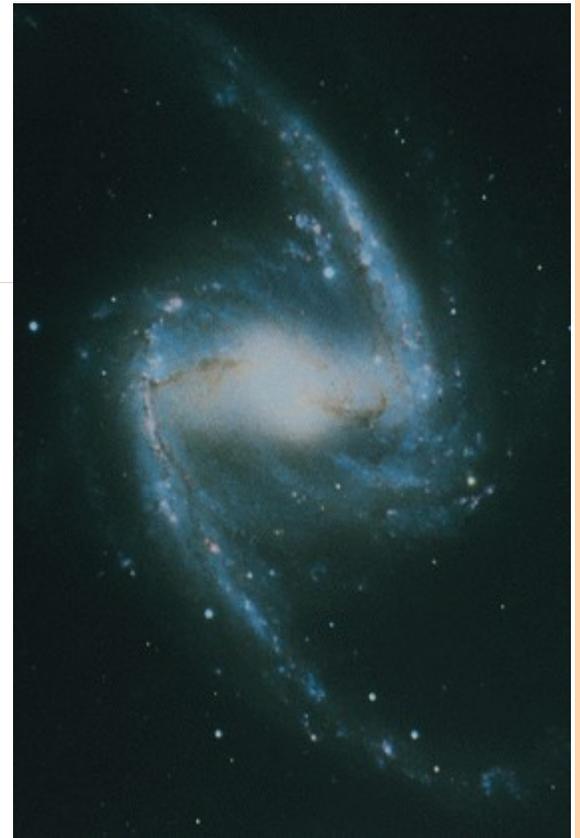


橢圓星系  
(elliptical)

不規則星系  
(irregulars)



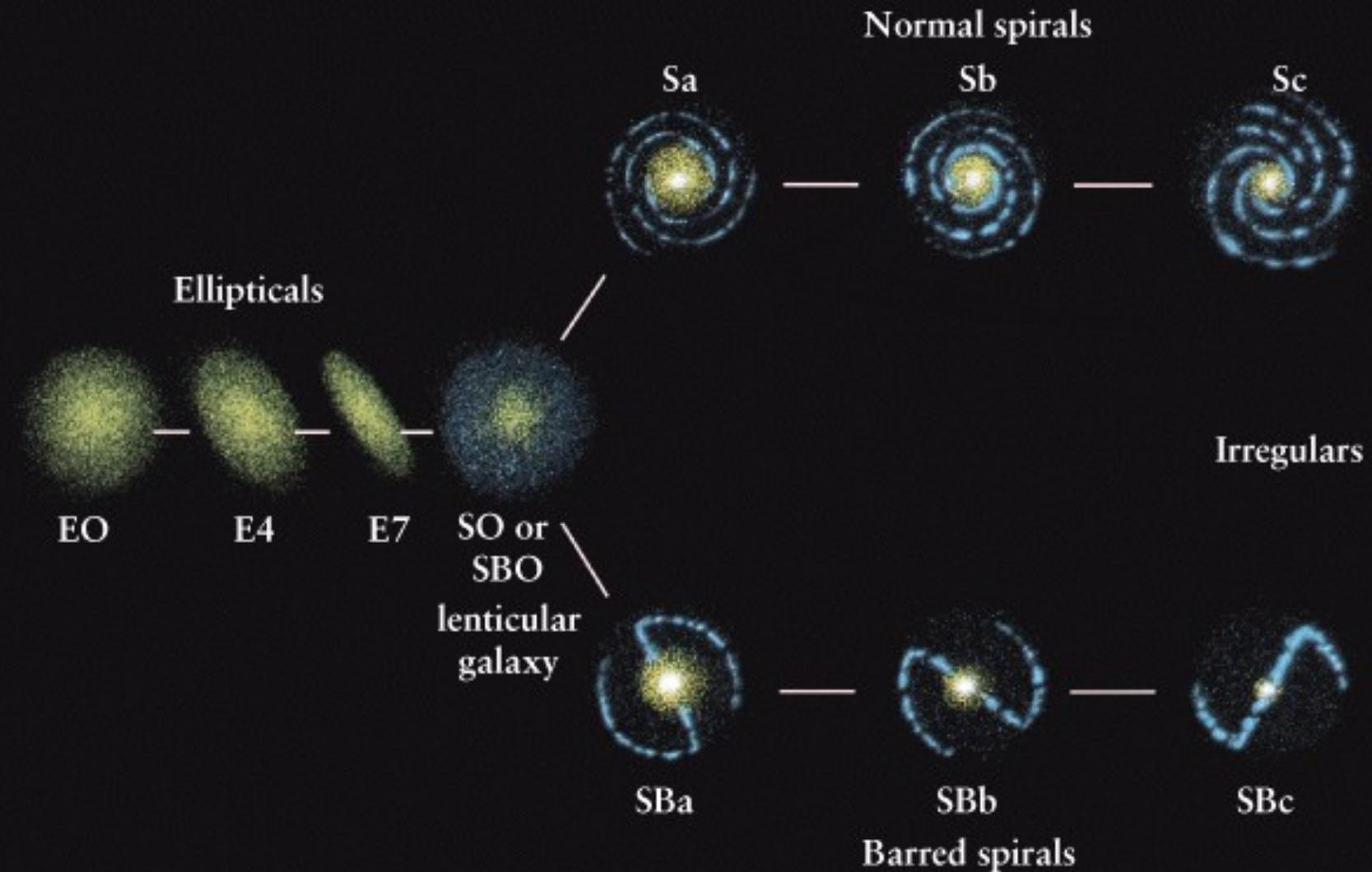
螺旋星系  
(spirals)



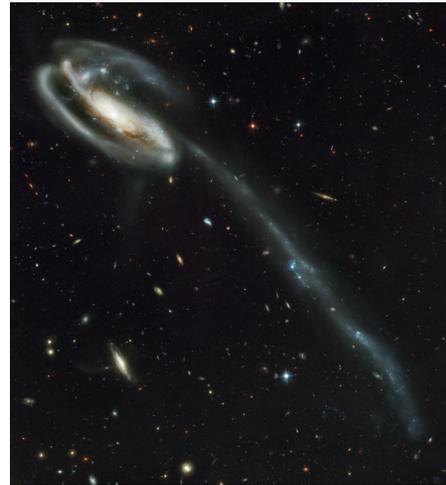
棒旋星系  
(barred spirals)

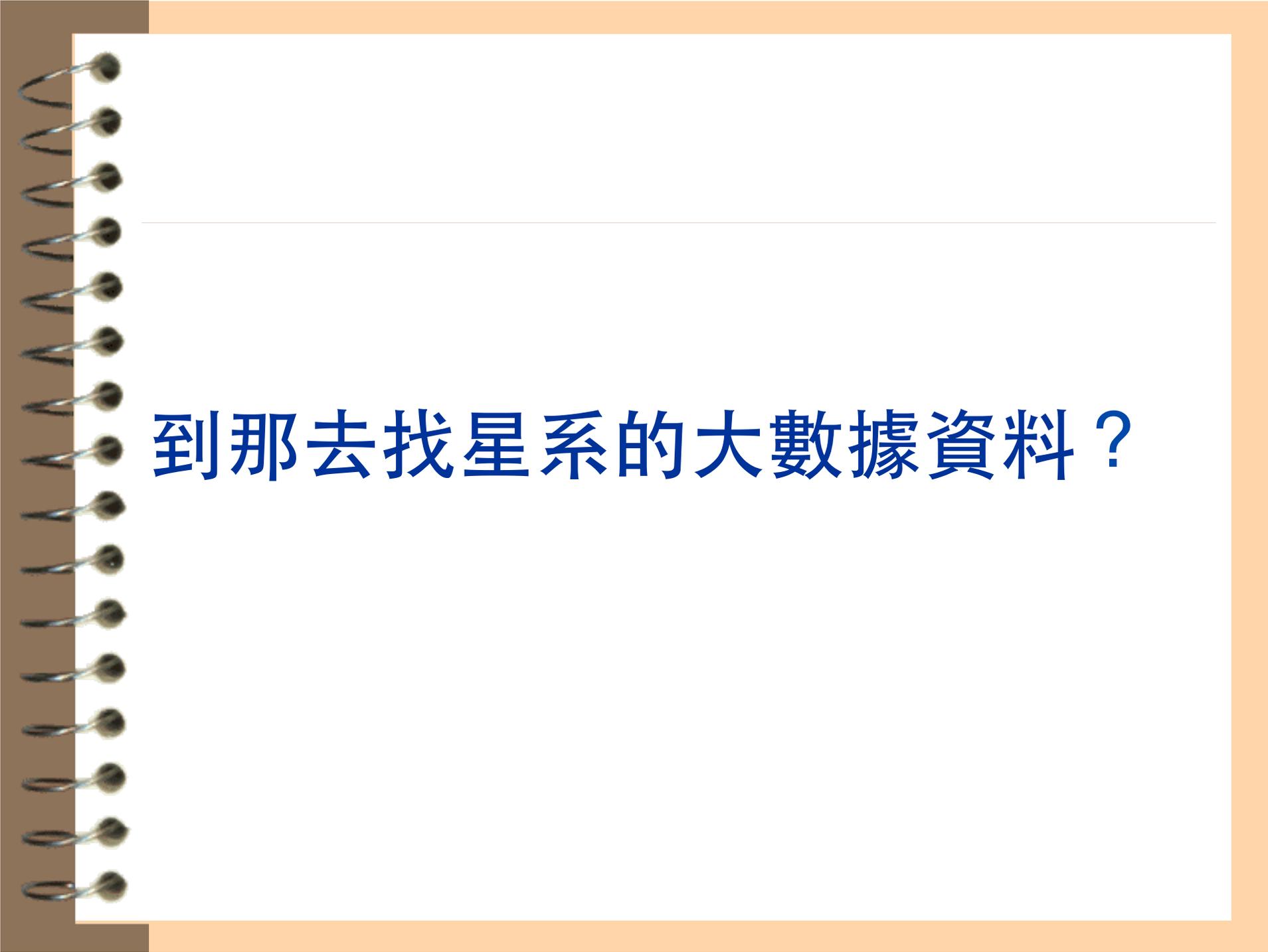


# 哈伯的星系音叉圖(Hubble tuning fork)



# 不規則星系



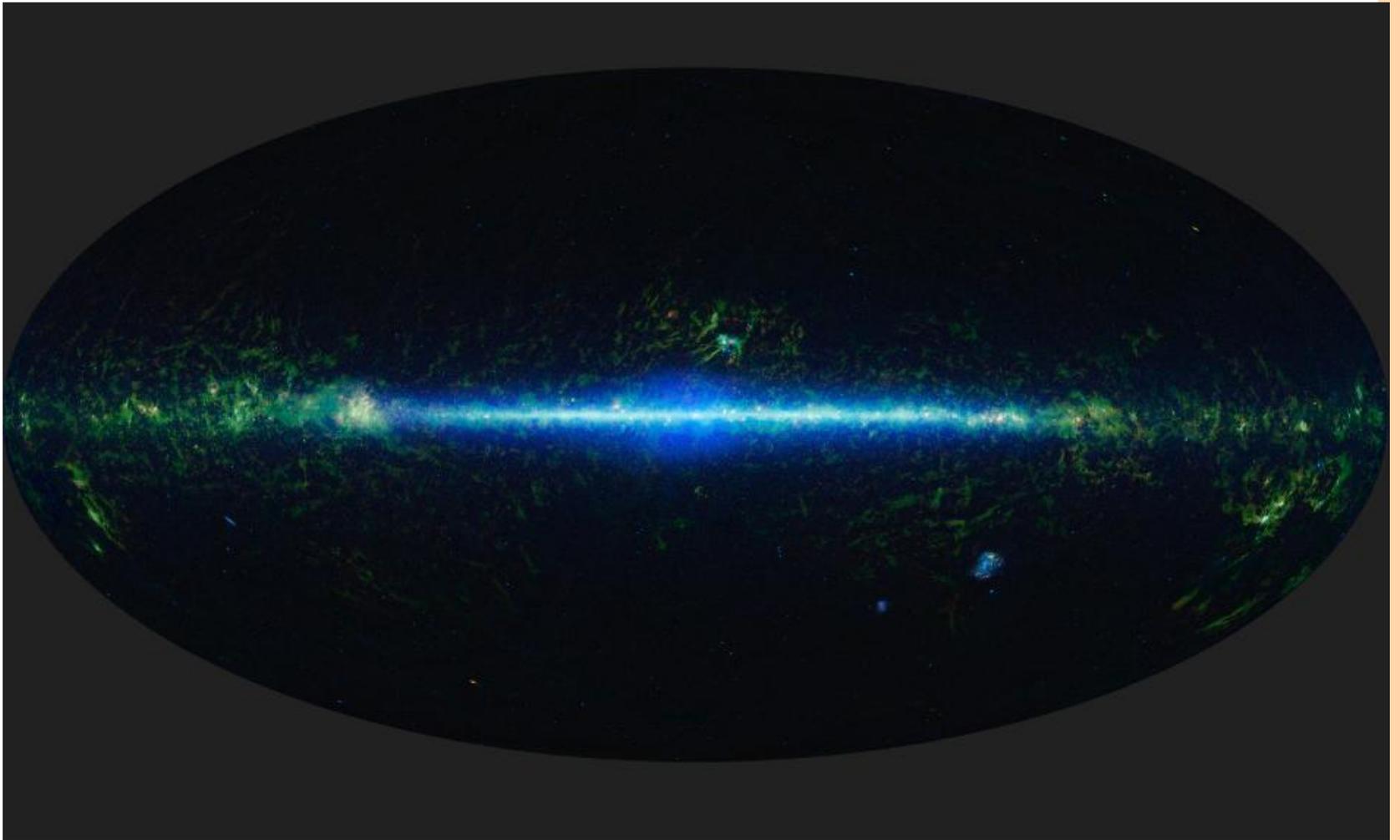
A graphic of a spiral-bound notebook with a brown cover and a white page. The spiral binding is on the left side. The page is mostly blank, with a horizontal line near the top. The text is centered on the page.

**到那去找星系的大數據資料？**

# Wide-field Infrared Survey Explorer (WISE)

- ✓ NASA的中紅外線衛星巡天計畫
- ✓ 共有四個波段3.4、4.6、12、22  $\mu\text{m}$
- ✓ 偵測到超過七億個點源和展源
- ✓ 其中約有一半以上是星系

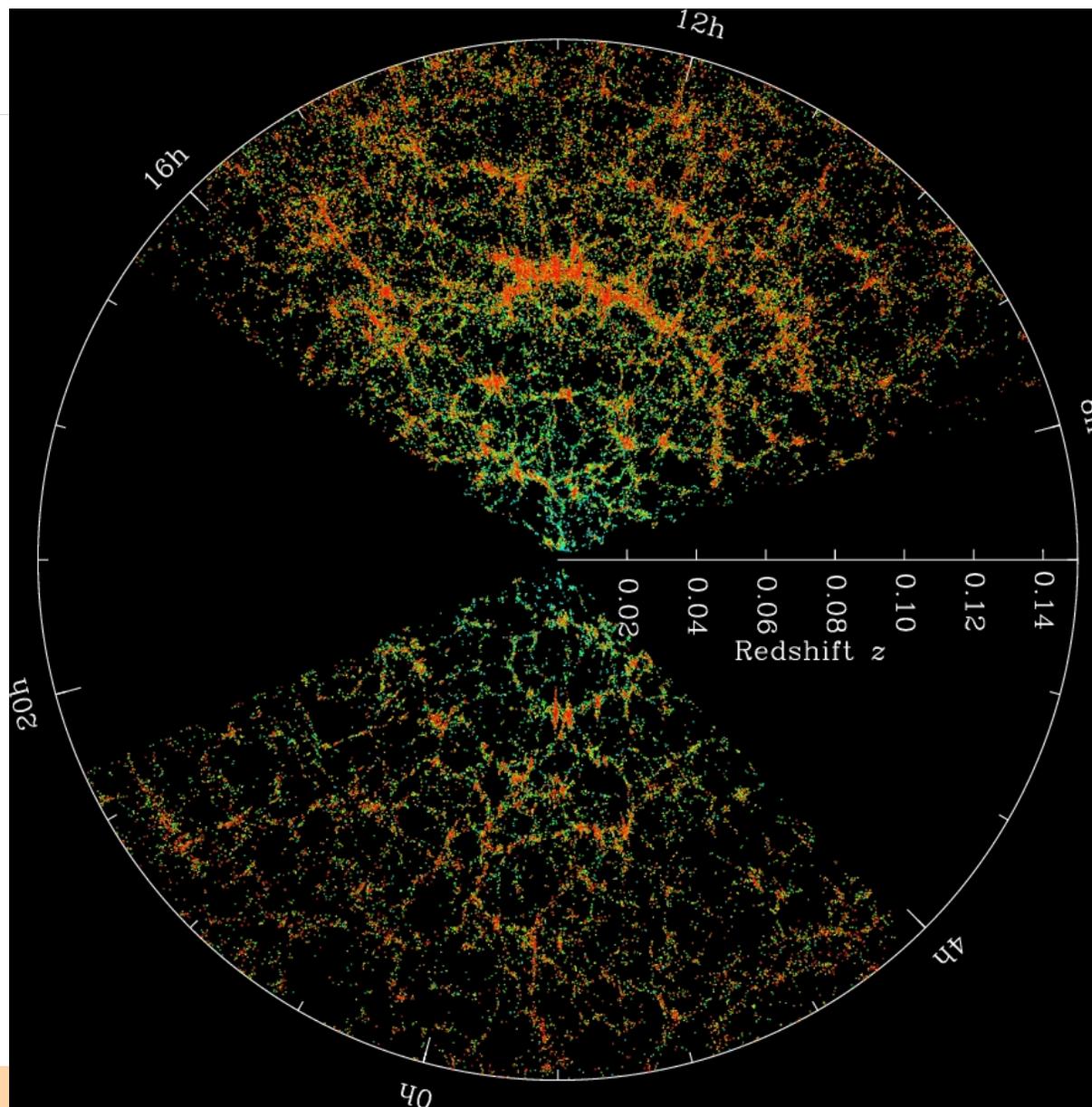
# WISE ALL SKY IMAGE



# Sloan Digital Sky Survey (SDSS)

- ✓ 可能是歷史上最成功的巡天計畫
- ✓ 五個可見光波段及光譜
- ✓ 偵測到超過四億個的天體
- ✓ 超過二億個星系
- ✓ 約一百五十萬個星系的光譜
- ✓ 三十五萬個類星體
- ✓ 每個天體的性質都有非常多的參數來描述

# SDSS所觀測到的宇宙大尺度結構



# 泛星計畫(Pan-STARRS)

- ✓ 「泛星計畫」(Pan-STARRS; the Panoramic Survey Telescope And Rapid Response System)。目的在搜尋太空中任何可能撞擊地球的天體，以及早做好因應對策。泛星計畫每晚將巡天6000平方度，深度達24星等，每月全天空巡天數次。泛星計畫特別適合探測「變化」(包括亮度與位置變化)的天體。泛星計畫的首要目標在標認出可能撞擊地球的小行星，其後果攸關人類文明的延續。但同時泛星計畫將產生極大資料量，涵蓋天空深度、廣度以及時間覆蓋。

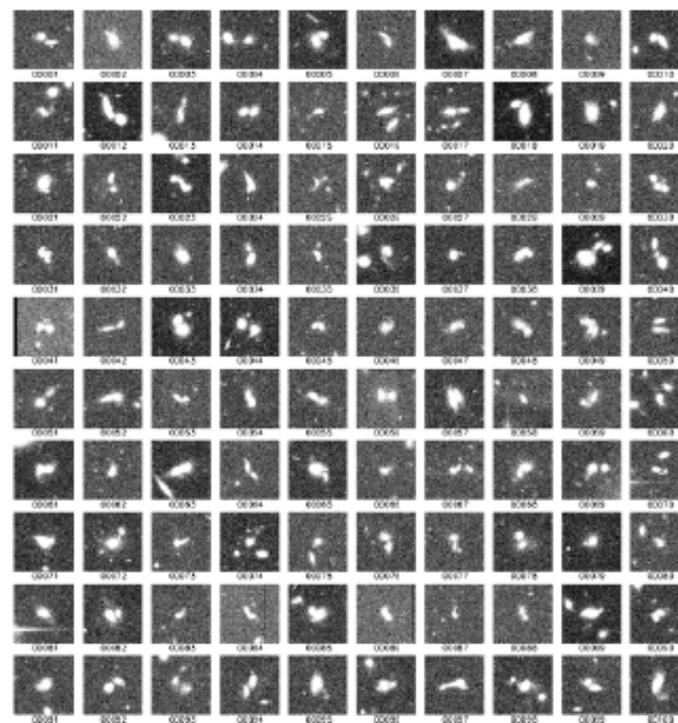
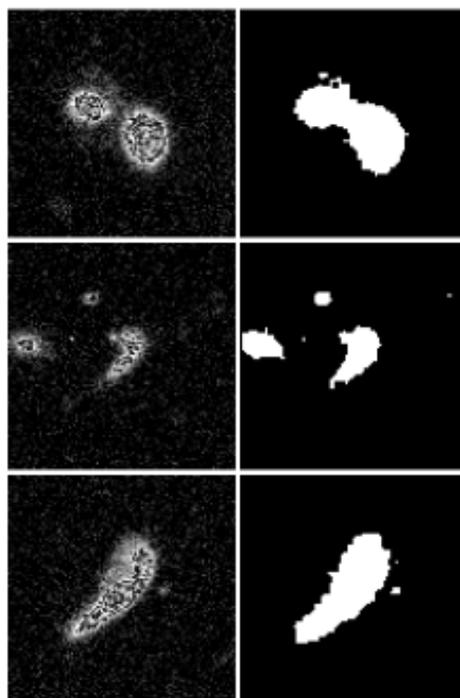
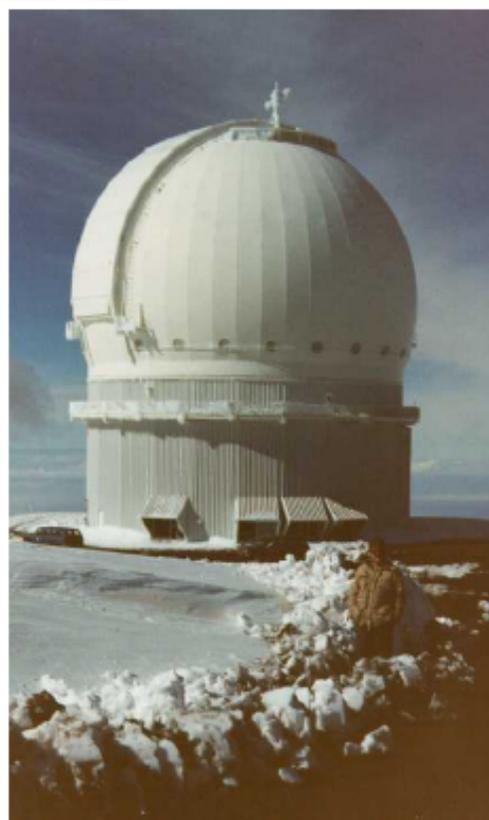
# 找尋危害地球的小行星及研究其它天文文現象



# 大數據星系天文學研究的例子

- ✓ 無法預測大數據的資料會產生什麼結果
- ✓ 以下純為個人偏見
  - 合併星系
  - 活躍星系核的統一模型
  - 紅類星體
  - 藍橢圓星系
  - 紅螺旋星系
  - .....

利用型態辨識法在CFHT所拍攝的四百多度的天區中所找到的一萬五千多個合併星系中的一百五十個.



Hwang and Chang (2009) ApJSS, 181, 233

# 非典型大數據方法

- ✓ 大數據科學的特徵之一是數據太大，而人力和人腦無法處理。
- ✓ 但某些能力，人腦是遠優於電腦
- ✓ 例如要從大的天文影像中指認星系的形態並不容易。
- ✓ 目前沒有可信賴的辨識程式，最可靠的方法是肉眼。

# Galaxy Zoo

---

- ✓ 一種很有名的方法是經由網路連結號召數十萬人共同組成志工。由眾人以肉眼尋找和分辨出各種星系型態。
- ✓ 非常成功的計畫

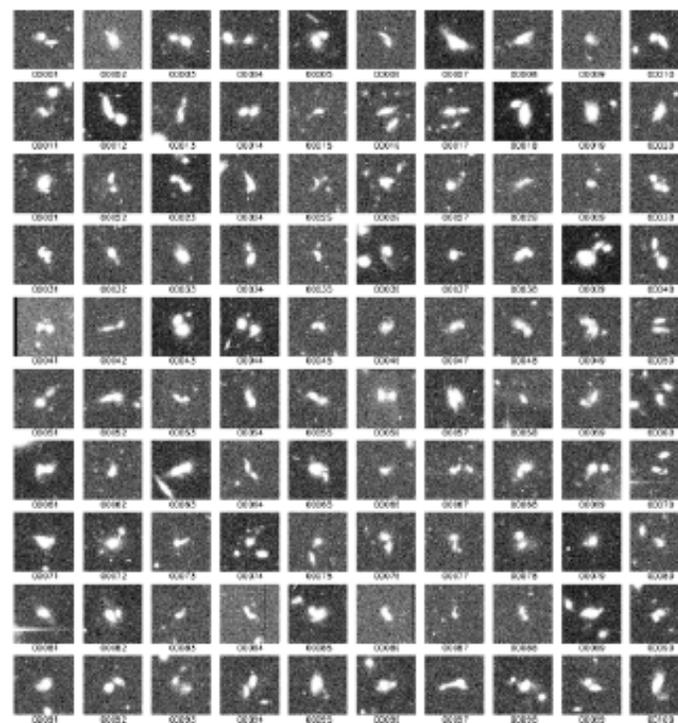
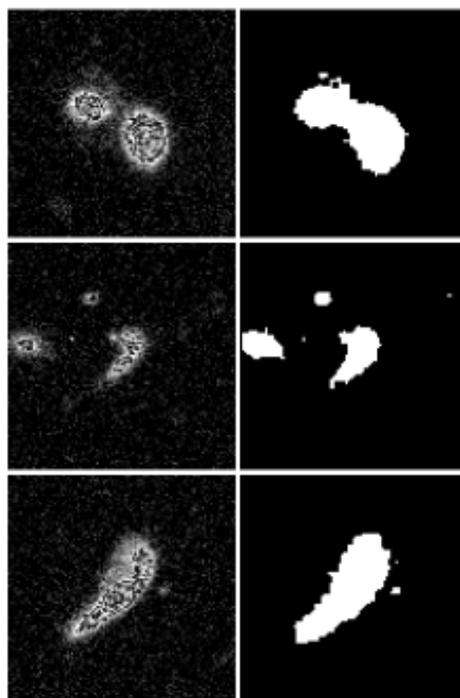
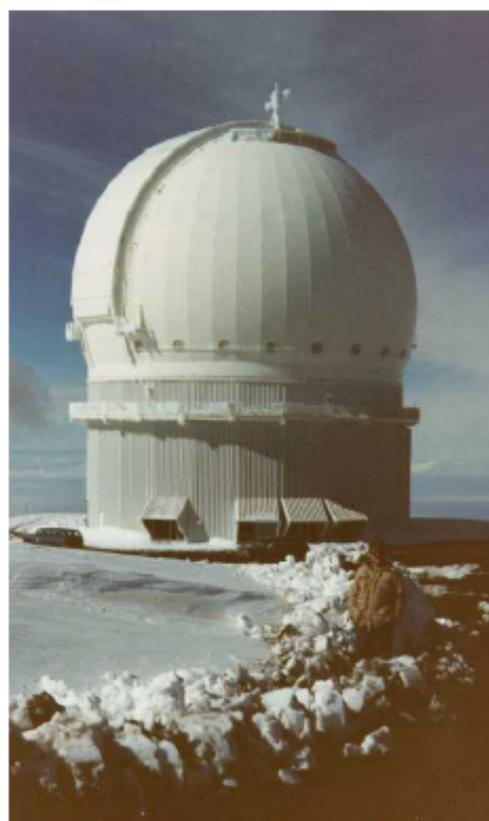
# 形態辨識法

- ✓ 也可以用形態辨識的程式來辨別星系形態
- ✓ 形態辨識的原理簡單的講就跟具有美肌模式修正的照片軟體的部份功能類似。
- ✓ 這些美肌軟體是要先找到人的臉的型態。
- ✓ 星系形態辨識的程式則要先找到星系的特別形態
- ✓ 在合併星系的運用比較成功。因為合併星系通常具有一些特別的形態，像是“橋”或是“尾巴”的形狀，可供程式辨識。

# 合併星系的形態辨識法

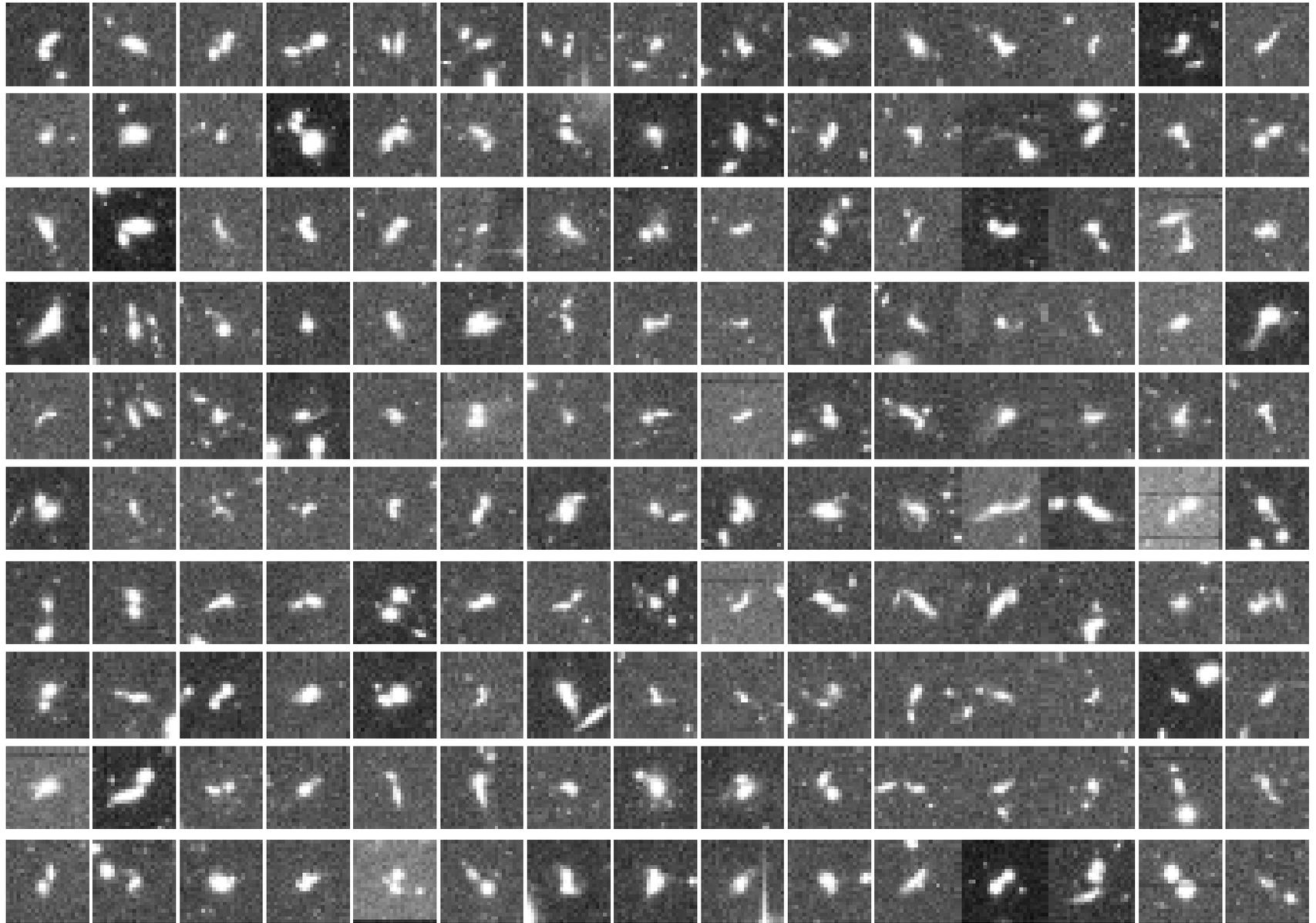
- ✓ Galaxy Zoo 團隊合併星系的結果發表在一篇2010的論文Darg et al. (2010)。這論文共有16位作者,經由約十六萬個志願者幫忙,搜尋超過一萬平方度的空間,共找到約三千個合併星系。
- ✓ 我的一位碩士生使用一種自行研發的型態辨識法並以肉眼為輔來尋找合併星系。在400平方度的空間中找到約一萬五千個合併星系。Hwang & Chang (2009)

利用型態辨識法在CFHT所拍攝的四百多度的天區中所找到的一萬五千多個合併星系中的一百個。



Hwang and Chang (2009) ApJSS, 181, 233

# 遙遠的宇宙還有更多的合併星系



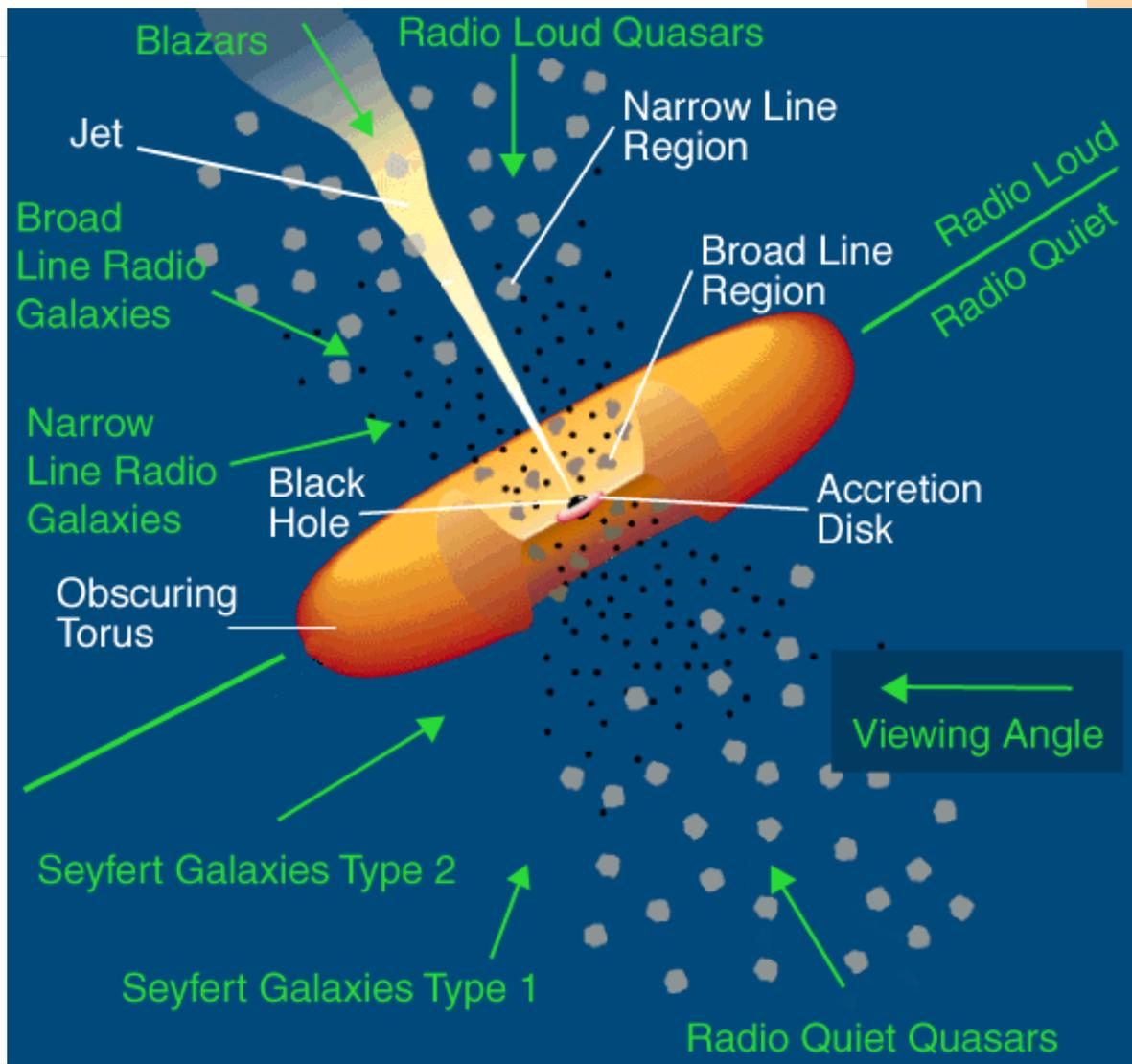
# 活躍星系核(AGN)

- ✓ 活躍星系核為宇宙中輻射能量最強的天體
- ✓ 能量主要來自該星系核心一個很小的中心
- ✓ 能量來源為星系中心超巨質量的黑洞
- ✓ 有許多不同類型的活躍星系核。
- ✓ 一般普遍接受的，活躍星系核統一模型，認為不同類型的活躍星系核主要是因我們的觀測角度不同所造成!

# 活躍星系核的統一模型

統一模型認為不同類型的活躍星系核是因為我們對天體的觀測角度不同所造成

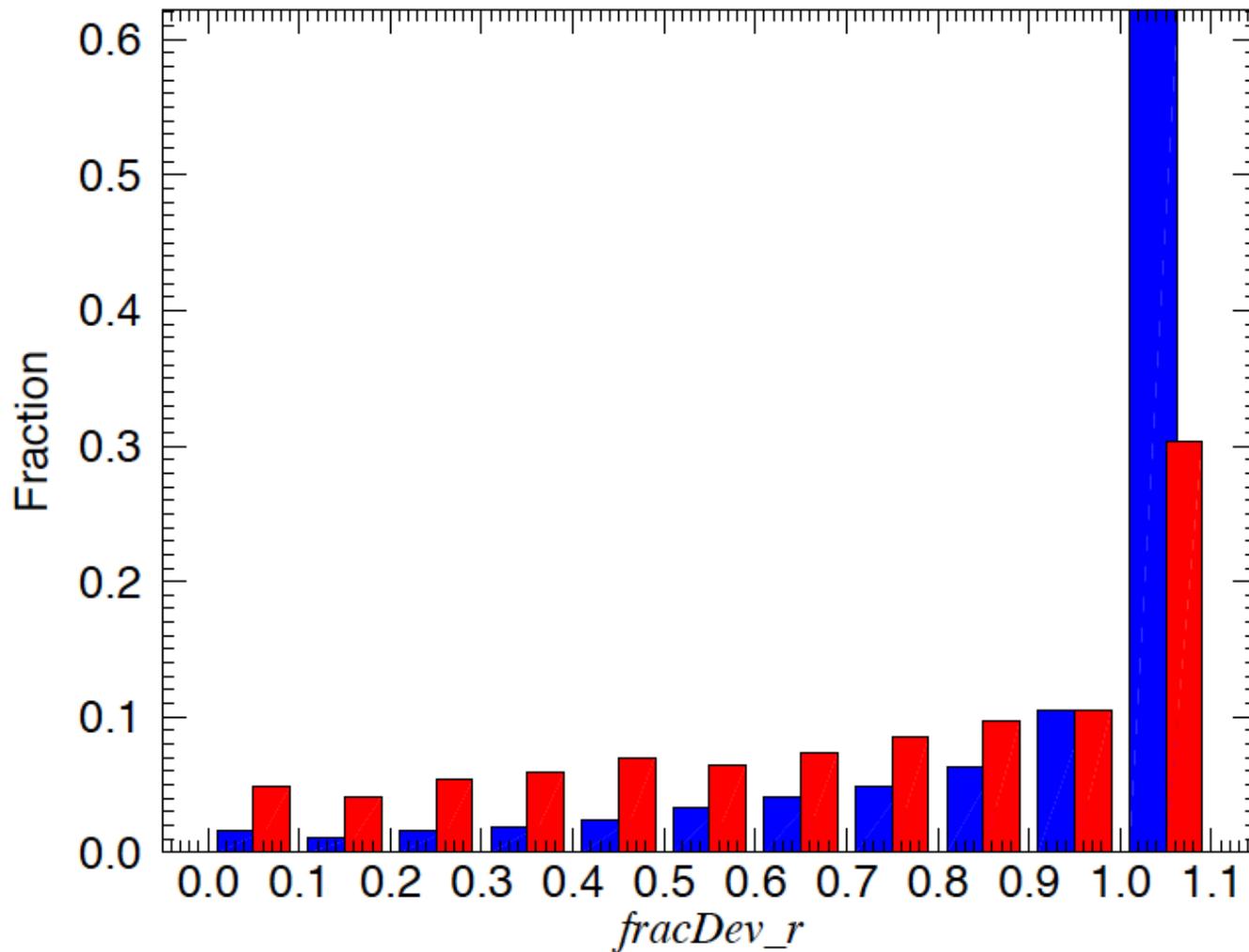
Credit: NASA



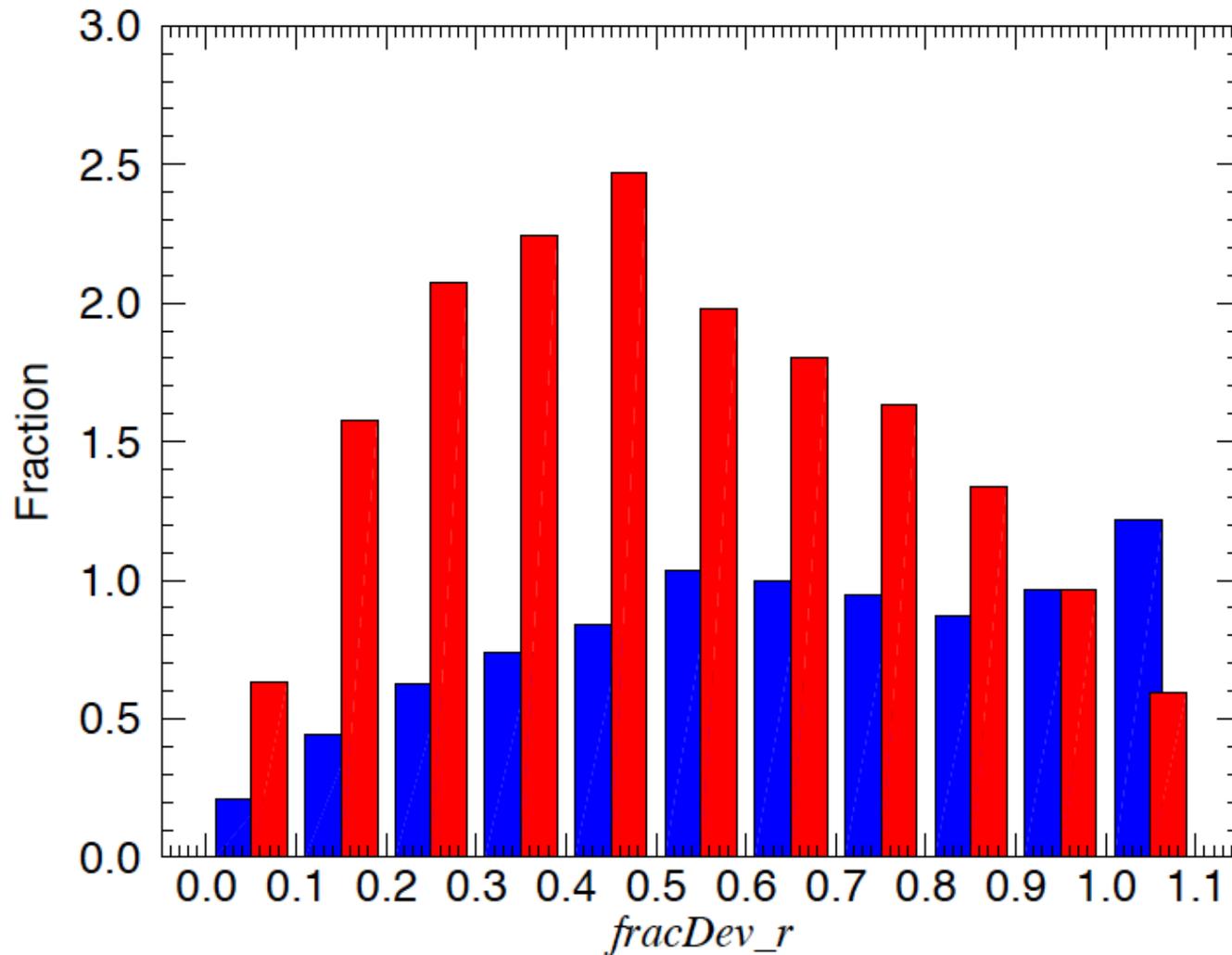
# 但是

- ✓ 我的一位學生利用大量數據的統計卻發現統一模型中的最佳範例，西佛一型和西佛二型活躍星系核的母星系有很大的不同分佈。
- ✓ 意謂著西佛一型和西佛二型差異不僅是來自我們觀測角度不同。它們的差異也跟它們的母星系性質有關！
- ✓ 統一模型並不完全正確！

## 二種西佛星系的母星系星凸比例



# 不同星凸下出現不同西佛星系的比率



A graphic of a spiral-bound notebook with a white page and an orange border. The spiral binding is on the left side. The text '謝謝' is centered on the page.

謝謝

# 思考題目一

- ✓ 上個月底有一則新聞說天蠍座的騎士最愛闖紅燈。台北市區監理所表示，駕籍在台北市的駕駛人104年1月到7月共有1萬3923件「機車闖紅燈」的交通違規案件，以違規駕駛人的星座進行比對分析，發現天蠍座（10月24日至11月21日出生）違規1282件最多（9.20%）；其次為摩羯座（12月21日至1月20日）1238件（8.89%）；而金牛座（4月20日至5月20日）1018件（7.31%）最少。
- ✓ 請問這統計的差異有效嗎？
- ✓ 要如何解釋這些結果？請用實際數據(可上網尋找)來陳述說明，而非僅用猜測或假設。

## 思考題目二

---

- ✓ Galaxy Zoo 的統計發現有相對較多的逆時鐘方向旋轉的螺旋星系
- ✓ 你認為可能的原因是什麼？
- ✓ 從已有的大數據資料，你要如何來驗證你的推測？