Galaxy Colour Profiles, using the Sloan Digital Sky Survey

Michael McEllin
November 11 2016

The object of this experiment is to determine whether different parts of galaxies differ in colour. The experiment can be repeated on different types of galaxy (e.g. spiral galaxies and elliptical galaxies), and we can also examine whether there are systematic difference in colour between different types of galaxy.

The main source of information will be the **Sloan Digital Sky Survey** (www.sdss.org) which has collected a superb set of images of that part of the sky which is visible from the Apache Point observatory in New Mexico in the USA. The sky was photographed in five different colours ranging from ultra-violet to infra-red (see voyages.sdss.org/filters). All the data is freely available to the public. It is generally referred to as the “SDSS”.

We will learn one method of getting hold of these multi-colour images, and how to process them using a freely available program written for educational use (SAOimage DS9 - see ds9.si.edu) to extract scientific results.

Equipment

A computer with:

- a standard web browser
- SAO Image DS9 - for image analysis
- 7-zip - a program for decompressing astronomical image files from the SDSS.

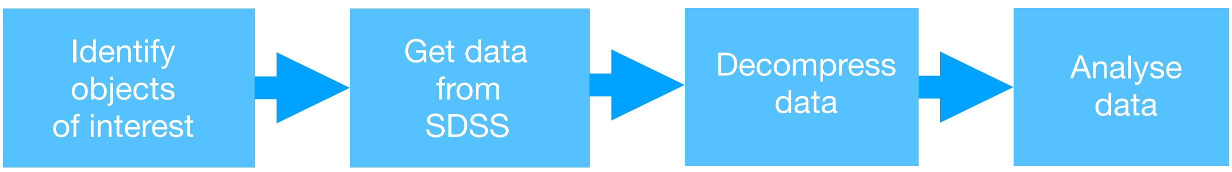
Prerequisite Knowledge

Basic astronomical knowledge:

- Some understanding of the electromagnetic spectrum (e.g. the range from ultraviolet to infra-red).
- What is a galaxy?
- The standard astronomical coordinate system for finding objects (e.g. Right ascension and Declination).

Process

We can break down the process into four basic tasks:



Step 1: Identifying Objects of Potential Interest

You will need to be able to find galaxies. Fortunately, many generations of astronomers have compiled catalogues of “nebulae” (which the early astronomers knew only as “cloudy” shapes that were distinct from stars).

The earliest catalogue of bright nebulae was compiled by Charles Messier, first published in 1771, though it has been modified at various times since then to correct original omissions and mistakes. All the objects in the Messier catalogue are identified by an “M” number, such as M31 or M33. It contains 110 objects. These are, of course, the most easily visible nebulae, so generally of interest to astronomers.

- Wikipedia has a useful list of Messier objects at https://en.wikipedia.org/wiki/List_of_Messier_objects.
- You can also see Hubble Space Telescope photos of 107 of these objects at https://www.nasa.gov/content/goddard/hubble-s-messier-catalog.

Some of the nebulae in these catalogues are now known to be galaxies. The rest are:

- Supernova remnants, for example the *Crab Nebula*, designated M1 in the Messier catalogue;
- Planetary nebulae, such as the *Ring Nebula*, M57;
- Star forming regions, for example the *Orion Nebula*, M42;
- Globular clusters, such as M2.

Of the 41 galaxies about 10 are spirals, sufficiently face-on to be useful for profiling, and with a sufficient proportion of their extent lying on a single SDSS *field*. (The SDSS survey is published as a set of fields of 2048 pixels square, which lie edge to edge to cover the sky. It is a matter of chance whether an object of interest is in the middle of a field, or lies across an edge. A few of the spiral galaxies, such as M31, the “Great Nebula in Andromeda”, have large angular extents and cover several SDSS fields. There are methods of stitching adjacent fields together to get more complete views, but these are more complex than we can cover at present.)

I suggest that you now take a look at one of: M51, M61, M74, M88, M90, M91, M94, M95, M99, M101. (Choose your own preference after looking at some images in SDSS- see below for instructions.) **Review this section for further choices when working on your CREST project.**

- Note down the object you have chosen, its type and position e.g. as in Wikipedia:

M101		Spiral	$14^h\ 03^m\ 12.6^s$	+54° 20′ 57″

The next most useful catalogue for our purposes is the New General Catalogue of Nebulae and Clusters of Stars (usually known as NGC) which has 7840 entries, including of course all the Messier objects, plus many fainter objects. Many objects imaged by the Hubble telescope come from this catalogue. For the purposes of a Silver CREST project there is, however, more than enough to keep you busy in the Messier catalogue.

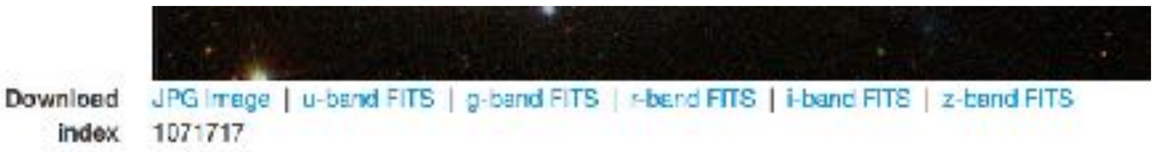
For your CREST project you may well want to look at a number of different types of galaxy, such as spiral, barred spiral, elliptical, “star-burst” galaxies, or even one with an *Active Galactic Nucleus*, such as M87 (though to have a 2 billion solar mass black hole at its middle).

We might at some point also want to look at the Abell Catalogue of clusters of galaxies. Abell cluster images sometimes show fifty or a hundred galaxies on the same SDSS field. All the galaxies in these clusters are relatively speaking almost at the same distance from us (and generally fairly far away and each of small angular extent on the image). This can be useful when looking at overall colour differences between different types of galaxy.

Get Data from the Sloan Digital Sky Survey

With your web browser, go to the SDSS “Sky-server” website at https://dr12.sdss.org/fields/ .
N.B. The instructions below work for the “Google Chrome” web browser and we recommend that you use that software. A slightly different procedure might be required for a different web browser.

- In the box on the top left labelled “**Search by Object Name**” type “M74” (without the quotation marks!) or any of the other Messier ids that you may select (e.g. M101) The Sky server will generally find objects that appear in most of the astronomical catalogues by their catalogue id. You can also enter right ascension and declination coordinates from a catalogue. (You may at this point wish to preview a number of Messier objects to find one that appears to have all the characteristics that you wish to see.)
- You should then see an image appear in the centre of the page. (You will also see the coordinates and the observing run number, the cameral column number and a field number. At present you can ignore these data.) Note that the coordinates are given in decimal degrees not the standard astronomical hours/minutes/seconds for right ascension or degrees/minutes/seconds for declination. This is often a more convenient convention to use with computers – but especially with right ascension the expression of the coordinate values look rather different. (The documentation claims that you can put hh:mm:ss etc. into the coordinate boxes to find an image frame. I have not tried.)
- Note that the galaxy M74 is not centred in the image. This often happens. It is possible, with a somewhat more complex procedure, to download all the images that relate to a certain area of the sky, centred on an object of particular interest, and then stitch them together into a larger mosaic in which our selected object is centred, but this process is more complex than we need to understand for our purposes.
- Immediately below the image there are a number of hyperlinks that allow you to download the JPEG colour image you see, plus monochrome “FITS” images taken through the five colour filters (as shown in the image below).
    - u is ultraviolet (the file is accessed via "u-band FITS”).
    - g is green (the file is accessed via "g-band FITS”).
    - r is red (the file is accessed via "r-band FITS”).
    - i is “near” infra-red (the file is accessed via "i-band FITS”).
    - z is infra-red. (the file is accessed via "z-band FITS”).



- Download all the files with a series of mouse click on each of the listed files. (You can save these files to a specific place with a right mouse click and selecting “*Save link as…*”.)
- I suggest that you create a special directory to hold your astronomical work and put all the downloaded material in there.
- FITS images are in a format known as “*Flexible Image Transport System*” widely used by professional astronomers because, unlike JPEG images, they contain <u>all</u> the pixel data recorded by the camera, and can also carry related data, such as calibration tables, at the same time.

- Open a File Explorer window on the director in which you have saved the FITS files.
- The FITS files you have downloaded will all have a "**bz2**" file extension, and this means that they are in the form of a "compressed archive" file.) They cannot be decompressed using the standard Windows archive extraction tools. You will need to use an application called **7-zip**, which has been specially installed on the Stroud High network for this project.
    - o There are a number of ways of using **7-zip** which are basically similar to the usual methods of decompressing files with standard Windows archive tools (should you be familiar with those). Use which ever method is most familiar to you.
    - o You should be able to invoke this with a right-mouse-click on the compressed file icon in Windows File Explorer and then selecting the **7-zip - "extract files"**.
    - o If you use "extract files", a subdirectory will be created containing the uncompressed file. You may wish to move all of the uncompressed files out of their subdirectories to a common directory for conveniences of finding them.
    - o If you use the "Open Archive" option a file explorer will appear showing the compressed archive contents, you will be able to drag the FITS file to your working directory (it will be decompressed as it moves).

You can now proceed to data analysis.

Data Analysis

If you have successfully downloaded images from the SDSS, you will now have one JPEG image, and five decompressed FITS images on your computer. Our aim now is to quantitatively compare the information in the FITS images (each taken through a separate colour filter).

For this you need to use the **SAOimage DS9** software (ds9.si.edu) which should have been installed on your computers.

Be warned! Although this software was designed for educational use, rather than use by professional astronomers, it is not by any means as highly polished as most of the commercial software with which you may be familiar. In particular the user guide and the reference manual are not at all as clear and informative as you might have come to expect, and seem to assume a fair amount of prior knowledge. (You can find some video tutorials on YouTube that might help - but I have not explored these. Remember that this is *free* software, developed by people whose real job is doing research in astronomy. Be thankful that they gave time freely to do what they did.) My instructions are based on a good deal of exploration and trial and error. They work, but there may be easier ways of achieving the ends for which I have aimed.

- Start-up DS9. Click on Windows icon (bottom left of the screen), then **All programs**, then **science**, then **physics** and then select '**SAOimage DS9**'
- Use the "**File**"->"**Open**" menu to select one of your FITS files. (Go for the "r" file first – there is likely to be more on the image.) N.B. some of the more frequently used menu options also appear as buttons in a bar above the image area. I leave you to explore the use of this shortcut.
- The image may look a bit pale with only the nucleus of the galaxy showing clearly - not at all like the impression you get from the JPEG image. There is, in fact, really a very large difference in surface brightness between the nucleus and the arms of the galaxy, and that is correctly reflected by the data in the FITS image file. This does not, however, display well on a computer screen because our eyes are not good at dealing with such large brightness differences. (The nice colour images of galaxies such as you see in the

JPEG image are actually somewhat manipulated for maximum visual appeal. These types of image are NOT used for scientific investigations.) We can, however, get a more useful visual impression by using the **Scale** menu.

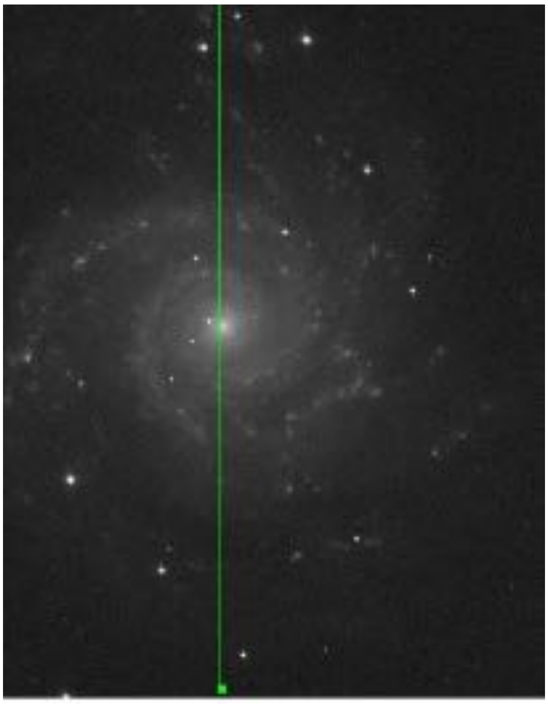
- Select "Scale"->"Log". You should see something that probably looks more familiar (though still in monochrome).
- In this way of looking at the data, each change by a factor of 10 in real surface brightness produces the same linear increment in the light on the screen. This has the effect of clearly showing the variations in brightness in the spiral arms, even though the absolute difference in brightness is small compared to the large difference between the spiral arms and the galactic nucleus.
- Note that we have not changed any of the original raw data, we are just displaying the data on a more convenient scale.
- DS9's **Scale** menu has a number of different options, for different purposes. The "**Log**" option seems to work well for our purposes.

- We now want to look at the profile of brightness along a line.
  - Select **"Region"->"Shape"->"line"**
  - On the **Edit** menu select "**Edit"->"Region**"
  - Use the mouse to draw a line across the galaxy from one edge to the other, crossing the centre. (Press the mouse left button and "drag" the line out.)
  - Make sure the line is "selected" – that is, click on it and be sure that you see dots at either end. (This can be a little fiddly on a line one pixel wide.) You can now use the "**Region**" menu to select **"Get Information"**. (You can also get the same result by double clicking on the line – again it is a little fiddly to select a one-pixel wide line.)
    - Note the coordinates at each end of the line, and the length of the line.
    - **ALSO**: Go to the "**Region**" menu and select **"Save Regions"** (down near the bottom). Make sure the **".reg"** file that is created by this action is in a sensible place on your computer, where you know how to find it. (For example, the same directory in which you are storing your FITS files.) This will save you time when you need to use the same line in a different colour image.
  - A new pop-up panel appears. On this panel, under the **"Analysis"** menu, select **"Plot2D"**. You now have a linear plot of brightness along the line.
  - On the plot window, use the **"File"** menu to **"Save Data"**. This creates a file of two columns with all the data required to do the plot. (E.g. you could load this file into Excel. There are more extensive graph plotting facilities within DS9 itself, but I assume that you will be more familiar with Excel, and would prefer to use this rather than learning a new tool.) It will be convenient to give this file a ".txt" file name extension when you name it. (See below.)
    - N.B. This is NOT a CSV file (i.e. easily opened by Excel in the same way you open an Excel workbook) the columns are separated by white-space so it must be *imported* into Excel using the Data Import tool, which can be told to recognise columns separated by white-space.
    - The instructions for importing data in columns into Excel differ a little depending exactly on which version of Excel you are using, but essentially follow the same logic. In my version of Excel: I click the "**Data**" item on the top level menu bar. This gives me a box labelled "*Get External Data*"

with an option “*From Text*”. A file chooser then opens and you must select the file containing the data from DS9. (By default, Excel may only display files with the “.txt” extension. Unless you have saved your data file with this extension you will have to use the “Display All Files” option.)

    - You will then be invited to confirm the way columns on each line are separated. Select “**Delimited**” (not fixed width), and then make sure the “**Tab**” and “**Space**” boxes are ticked.

  - The image below has logarithmic scaling of the brightness and shows the line that I choose to use on M74:

  - 

- We now want to do a plot along *exactly* the same line for a different colour (say green). This could be tricky if we try to set up exactly the same line with the mouse. We can, however, manage this quite easily using facilities in DS9:

  - **Before you load the green data** if you have not done so already, make sure that you have used the “**Region**” menu (as describe earlier) and select **“Save Regions”** (down near the bottom) to save a description of this plotting line. Make sure the **“.reg”** file that is created by this action is in a sensible place on your computer, where you know how to find this file. (For example, the same directory in which you are storing your FITS files.)

  - Now load the “green” (g) image file using the “**File**” menu.

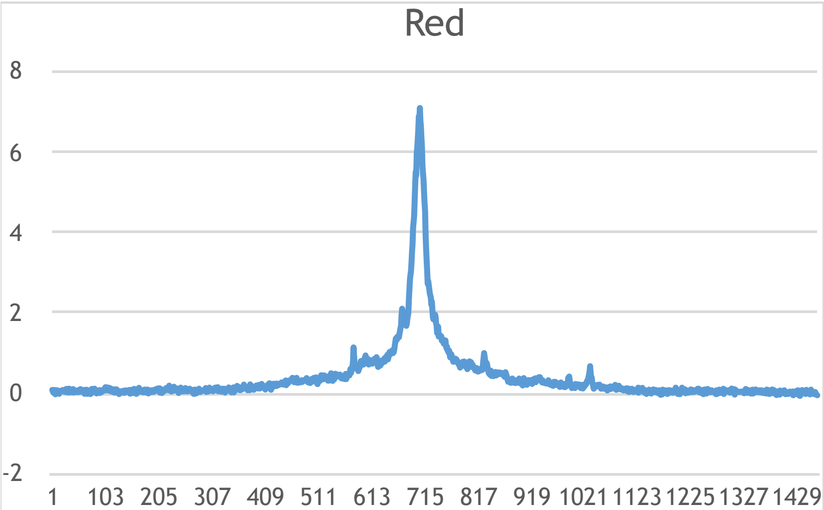
  - Use the **“Region”->”Load Regions”** option to load the .reg file you created on the previous set. (Click the “**Load into All Frames**” option when it comes up – though this will not matter to us now, you may find it convenient in future work.)

  - The same line now appears on the image and we can plot the green data along this line in the same way as with red data. Save the data to a file in the same way. Annotate the file name to show which colour is being saved. (**Be sure that you do not overwrite the file of red data by just accepting the default file name again!)**

  - We now have two files of data that can be loaded to Excel and we can compare them in various ways. I leave it to you to devise informative ways of comparing the data. You may, for example, wish to generate additional columns in the spreadsheet, using arithmetical operations (e.g. subtracting one column from another, dividing one column by another). What would be a sensible way to compare differences between red and green that might only show up in certain parts of the image?

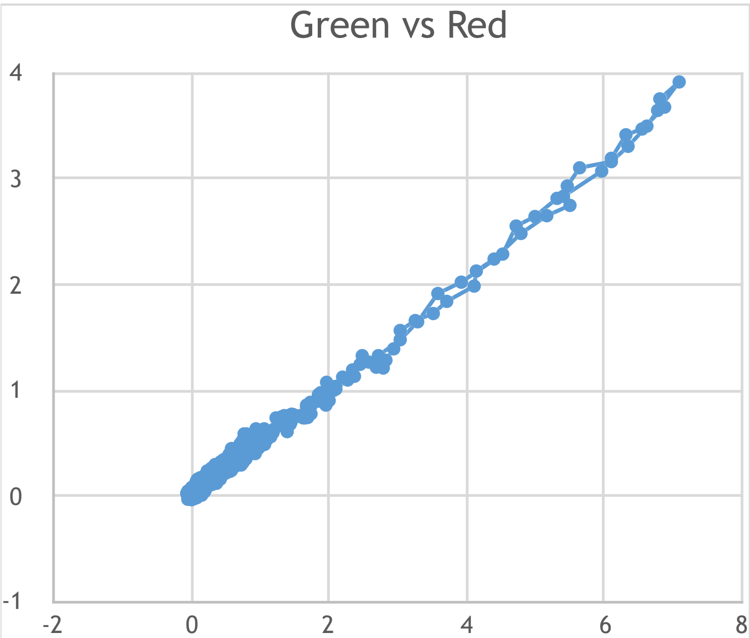
- Note that with the default graph plots the axis scales are automatic, and the shapes will look similar, though in fact there are higher peaks on the "red" graph in absolute terms. You will see this when you import both datasets to Excel.
  - o You may like to try plotting both curves on the same graph.

The figure below, for example, is what I get when I plot the profile for the "red" filter FITS file.



Note the very large peak. Visually, we do not get this impression. Outside the core, the galaxy surface brightness is very low indeed.

If I plot green against red, using Excel, I get the following:



Is this a sensible way of comparing the red and green plots? Why? (Or why not?)

At this point I suggest that you do NOT do as I have done above, but try some alternative, and possibly better ways of analysing the data.

- Rather than using a line running from edge to edge, I suggest that you draw several lines from centre to the edge of the galaxy. (This avoids two lines lying on top of each other in the Green vs Red graph above.)
- You will need to look at comparisons not just between red and green, but between the other colours (e.g. red and ultra-violet, or red and infra-red). There is interesting information in all of these comparisons. There may be better ways of bring out any colour differences (e.g. as ratios, or as, for example (red-green)/(red+green). You need to experiment.
- You will probably find that the differences in colour as you cross the spiral arms look much less impressive in the graph than it appears in the JPEG image. This may be partly due to image manipulation for visual effect. However, if you look closely at the images you may well find that the really blue regions of the spiral arms are <u>not</u> continuous, but confined to small areas of star formation. We need to investigate this. The question is whether the background colour in the spiral arms is the same as in the nuclear - except for the small regions of star formation. (See “Getting Statistics for a 2D Region” below.)
- There are a number of possible investigations you may like to pursue (you will probably not have time for all of them):
  - Are galaxies with open spiral arms different to those with tightly wound arms?
  - Are “barred” spirals different to other spirals?
  - You may like to investigate the difference in profiles between spiral galaxies and elliptical galaxies.
    - Are elliptical galaxies generally redder than spirals?
    - Is the nucleus of a spiral galaxy pretty much the same colour as the overall colour of an elliptical galaxy?
    - Is there any correlation between the overall brightness of a galaxy and its size? (Look for an <u>Abell cluster</u> with 40+ members, such as Abell 426 - type this into the Sky server *Search, Object Name f*ield.) You could then use the techniques of the next section to work out relative colour differences and overall brightnesses.

Getting Statistics for a 2D Region

It is often interesting to find out how much light is within a designated region of the image. (For example, we might need to do this to handle one of the questions posed above.) This is one way of dealing with that question.

- Go to the **Region** menu and select **Shape—>Circle.**
- Now use the mouse to drag out a circle around an area of the image. (That is, depress left mouse button and move the mouse while holding it down.)
- Make sure that the region is “selected”. (I.e. click on the line defining the region and be sure that four dots sit at its corners.
- In the Region menu, select **Region—>Get Information**. A new window appears. (You can also get this by a double click on the region boundary.)
- Within the new window, select **Analysis—>Statistics**. A new window appears in which various numbers characterising the region are printed. (Depending on which font your computer uses for display you may find that the number do not line up under their proper headings, so be warned! Count fields from the left. Note that you can change the default font in DS9 for

better alignment. Still be warned! Sometimes two adjacent numbers run into each other.) There are several numbers of real interest for our purposes:

- **Sum**: the total amount of light in the region.
- **Area**: the total area in arc-seconds squared.
- **Surface brightness**: the amount of light per square arc-second. This is a good number for comparisons between different regions of the image, proving the regions are positioned so they discriminate correctly between different types of area.
- You should copy and paste regions to ensure that you are using regions of exactly the same dimensions to measure different parts of the image. Or, you can use the **Get Information** window to specify shape exactly with numbers.

As previously, you can **Save** the region definitions and use the same definitions on images in the different colour filters.

I have not told you exactly what the numbers you are plotting actually mean in scientific terms. In fact, this is not necessary as we are really only interested ratios, so the units do not matter. There is a way of relating these numbers to the standard astronomical "magnitude" scale, but it is rather complex and involves careful calibration (a typical scientific concern in most astronomical observations). It might, however, just amuse you to know that the numbers you see are actually in units of "*nanomaggies*".

Limitations of SAOImage DS9

There are several things that we might like to do with our images that are difficult or impossible with DS9. Other tools let you, for example, add, subtract and divide entire images, pixel by pixel. In my judgement DS9 was the best compromise between usability and functionality. **SalsaJ** (also designed for educational use) is good for adding/subtracting/dividing entire images, and also plots profiles, but does not first align the images using the astronomical coordinate system (so, the profile line can be in a slightly different place on each image). Professional astronomers use more more complex tools with more functionality, so they take longer to master, but they still need to switch between different software tools for different types of analysis. It is part of the trade skills.

Final Remarks

We have explored only one way of getting one type of data from the Sloan Digital Sky Survey, and one way of using that data. There are many types of data in the SDSS, many ways to get data from SDSS and many ways of using it. Hundreds of professional astronomers use the SDDS data every day for cutting edge research in astronomy. In six, seven or eight years time some of you may be doing undergraduate projects, or signing on for astronomy PhDs in which you will work with the same collections of data.

Nobody is going to stop you exploring what you can do with this data now, but you will have to put in some work. A good place to start further exploration is the Education section http://www.sdss.org/education/ and particularly the "Voyages" area http://voyages.sdss.org which has a number of tutorials at a variety of levels. I shall be looking at this very carefully in the near future to see if I can devise some "Gold" level CREST projects, and I hope to see some of you doing those next year.

You may also like to look at websites such as Microsoft's World Wide Telescope (worldwidetelescope.org) which also uses data from the SDSS (and many other astronomical databases for different wavebands from microwaves to X-rays). This can be a good way of exploring the sky to find interesting objects.

The World's astronomers have agreed that essentially *all* astronomical data should now be made freely available via the World Wide Web. Much of it is already there, but you have to learn different methods of getting hold of it. (For example, every image captured by Hubble is available in the format used for scientific analysis by professional astronomers - if you know where to look.)

It would, however, be better if all this data were accessible in a standard way that will work for visual, radio, x-ray and gamma ray data from all of the World's telescopes and satellites. This is known as the "Virtual Observatory" project, and I think that it is a very exciting prospect.