

Un caso riuscito di contaminazione scientifica tra ambienti diversi

Giuseppe Di Persio

Istituto Nazionale di Astrofisica

Abstract. Viene brevemente descritta la genesi del formato FITS, nato in ambiente astronomico, la sua evoluzione e come sia divenuto parte integrante del progetto di digitalizzazione dei manoscritti della Biblioteca Apostolica Vaticana.

Keywords. Conservazione Digitale, Biblioteche, Astronomia

Introduzione

Si spiegherà come un formato usato solo in Astronomia sia approdato e sia stato usato dalla Biblioteca Vaticana nel suo progetto di conservazione digitale.

1. Nasce un problema, nasce il FITS

Negli anni 70 ci fu un grande sviluppo dell'Astronomia e dell'Astrofisica. I sempre più numerosi satelliti in orbita, i telescopi e i radiotelescopi sulla Terra produssero una enorme (per l'epoca) quantità di dati. Questo fece sorgere alcuni problemi, il principale dei quali era la pochissima standardizzazione nei formati di salvataggio, fattore che ostacolava molto la diffusione del sapere e lo studio condiviso.

La comunità scientifica astronomica raccolse la sfida, studiò il problema e nel 1979 furono rese pubbliche le specifiche di un nuovo formato di salvataggio dei dati: era nato il FITS (Flexible Image Transport System), standard dal 1981 e che divenne ben presto il formato più usato per la memorizzazione di immagini e dati in Astronomia. Fino al 1987 questo formato era gestito direttamente dalla NASA tramite un suo dipartimento (il NASA Office of Standard and Technology), poi la gestione passò allo IAU (International Astronomical Union), la più importante associazione di scienziati del settore.

2. Cambiare rimanendo se stessi

Il FITS è usatissimo ancora oggi a quasi quarant'anni dalla sua nascita, un periodo che in informatica equivale a un'era geologica. I motivi di questo successo sono da ricercarsi nel fatto che è stato creato tenendo ben presente il problema della conservazione digitale a lungo termine, quindi è semplice ed efficiente.

Le sue caratteristiche principali sono:

- è esente da royalty e di libero uso,

- tutte le informazioni necessarie a leggere, capire e rappresentare i dati contenuti sono facilmente leggibili e accessibili,
- la sua ampia diffusione e uso sono garanzia di longevità,
- è flessibile, potendo salvare molti tipi di dati con una grande quantità di metadati, ma robusto relativamente alla possibile perdita di informazioni,
- esiste un'ampia documentazione sul formato tutta di pubblico dominio,
- è completamente indipendente dal sistema operativo in uso,
- è personalizzabile senza perdere l'aderenza allo standard,
- gestisce senza problemi file di grandi dimensioni.

Un file FITS è composto da due elementi distinti, che si possono ripetere più volte. Il primo elemento è l'HEADER, formato da testo ASCII visualizzabile facilmente in ogni ambiente operativo, mentre il secondo elemento è composto dai dati in formato binario. Nell'HEADER ci sono delle "Keyword" che definiscono dei valori per una serie di parametri: alcuni (pochi) obbligatori, altri facoltativi che servono ad inserire informazioni aggiuntive. Questo schema HEADER-DATA si può ripetere più volte e permette di memorizzare nello stesso file più immagini o dati correlati.

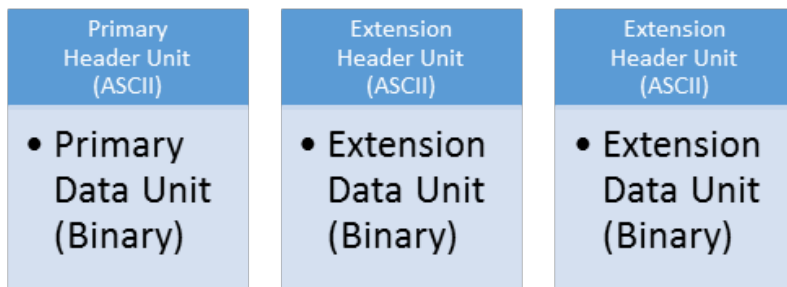


Fig. 1
Struttura di un file FITS

Esempio di HEADER di un file FITS:

Keyword Obbligatorie:

SIMPLE = T / Il file è conforme allo standard

BITPIX = 16 / Numero di bit per pixel

NAXIS = 2 / Numero degli assi

NAXIS1 = 1024 / Risoluzione dell'asse 1

NAXIS2 = 1024 / Risoluzione dell'asse 2

Keyword facoltative:

AUTHOR = ' JOHN SMITH'

DATE = '05/06/1984'

OBJECT = ' True photos of aliens on Mars'

ORIGIN = ' ASA secret archive'

DATAMAX = 1.246536493E+05 /MAX PIXEL VALUE

DATAMIN = 1.483317930E+01 /MIN PIXEL VALUE

COMMENT = ' Esempio di un possibile HEADER'

HISTORY -----

HISTORY Permission is granted for publication and reproduction of this
HISTORY material for scholarly, educational, and private non-commercial
HISTORY use.

HISTORY NRAO
HISTORY 520 Edgemont Road
HISTORY Charlottesville, VA 22903-2475
HISTORY -----
END

Attualmente gran parte dei dati astronomici ed astrofisici sono memorizzati in FITS. Lo IAU cura l'evoluzione del formato, giunto alla versione 4.0 nel luglio 2016, in modo da mantenerlo sempre in grado di rispondere alle nuove esigenze della scienza, ed è prevista una articolata struttura basata su comitati locali per raccogliere e discutere le richieste di cambiamento che arrivano dalla comunità scientifica internazionale.

Ogni cambiamento introdotto ha arricchito il formato di nuove funzionalità ma ha sempre mantenuto la piena compatibilità con le precedenti versioni. Il motto ufficiale del formato: "Once FITS, always FITS", finora non è mai venuto meno!

3. Il FITS e la Biblioteca Apostolica Vaticana (BAV)

La BAV, tra le più importanti biblioteche al mondo, da oltre 500 anni istituzionalmente svolge un ruolo non solo di diffusione ma anche di conservazione, protezione e restauro del patrimonio librario custodito, soggetto a continua usura anche tramite il semplice contatto con la pelle umana.

Per conservare i suoi manoscritti fin dal 2007 fu avviato uno studio di fattibilità per comprendere come le tecnologie informatiche potessero aiutare a risolvere il problema, e la quasi scontata conclusione fu quella di digitalizzare il suo intero patrimonio librario. Naturalmente la digitalizzazione doveva essere ad altissima risoluzione, con una fedeltà assoluta dei colori e con l'obiettivo, ambizioso, di rendere fruibili i files dei manoscritti almeno per i prossimi 100 anni.

Se si pensa alla rapida obsolescenza di quasi tutto quello che ha a che fare con l'informatica si capisce come il lavoro non sia stato per niente facile. I pilastri del progetto di digitalizzazione dovevano essere il formato di memorizzazione e una serie di procedure volte ad evitare l'obsolescenza tecnologica dei sistemi informatici.

Mentre per il secondo punto sono state stabilite alcune stringenti linee guida e presi accordi con le aziende fornitrici dell'hardware per periodici aggiornamenti almeno biennali dei sistemi, per quanto riguarda il primo punto sono stati fatti attenti studi comparativi tra i vari formati esistenti e presa in considerazione la possibilità di crearne uno nuovo.

Purtroppo, nessuno dei formati tra quelli maggiormente diffusi in ambito fotografico aveva i requisiti necessari. Neanche quello finora più usato, il TIFF, perché:

- è gestito da una compagnia commerciale privata, e non è escluso che in futuro siano esigibili pagamenti di royalties
- non gestisce file maggiori di 4 GB mentre già esistono scanner che producono file più

grandi

- non lavora a 64 bit e non permette di salvare tutta una serie di importanti informazioni legate alle immagini, come quelle storiche o bibliografiche.

Alle prese con questo problema e venuta a conoscenza del FITS la BAV ne intuì presto le potenzialità, anche se non era mai stato usato al di fuori dell'ambito astronomico, e iniziò una serie di test.

Vista la provenienza del FITS la Biblioteca si avvalse fin dall'inizio della collaborazione di esperti dell'Istituto Nazionale di Astrofisica, e iniziò così una sinergia tra settori apparentemente diversi accumulati dai problemi della Long Term Data Preservation.

Questa collaborazione si è rivelata estremamente proficua e ha interessato anche altre importanti istituzioni del settore spaziale come lo IAU, che ha attivamente collaborato per arricchire il FITS con delle keyword specifiche per il settore della conservazione digitale, o l'ESA (European Space Agency), da sempre interessata alle problematiche della LTDP, che ha firmato con la BAV degli accordi per condurre studi congiunti nel settore.

Dopo una serie positiva di test la BAV ha fatto partire il progetto di digitalizzazione dei manoscritti.

I numeri del progetto:

- • 80.000 manoscritti
- • 150.000 stampe, mappe e disegni
- • 9.000 incunaboli
- • +130.000 unità archivistiche
- • Tutti scansionati ad altissima risoluzione
- • Oltre 40 PB di dati originali

Il progetto, che ha suscitato l'interesse di altre istituzioni del settore, è basato sull'uso di FITS per salvare le immagini e su XML per tutte le informazioni bibliografiche o di altra natura non gestibili da quel formato (il FITS salva dati in formato ASCII e/o binario, e non permette di salvare testo in caratteri non ASCII). Al momento è stato digitalizzato poco più del 20% del patrimonio librario della Biblioteca.

Conclusioni

La digitalizzazione dei manoscritti è stata non un punto di arrivo ma di partenza per una serie di ulteriori interessanti sviluppi.

In generale solo una frazione del patrimonio librario conservato nelle grandi Biblioteche è attentamente studiata, essenzialmente a causa della delicatezza dei manoscritti e del fatto che devono essere studiati di persona. La digitalizzazione di altissima qualità in corso permetterà agli studiosi di analizzare i testi senza dover spostare o toccare i manoscritti, operazione facilitata anche dalla partecipazione della BAV come socio fondatore al Consorzio IIF e quindi dalla possibilità di usare questo interessante standard per la visualizzazione in alta definizione degli elementi digitalizzati.

Inoltre si potranno usare tecniche automatiche di riconoscimento del testo o di immagini che permetteranno indagini e studi su scala mai tentata prima.

Un ultimo ma credo significativo risultato è che la BAV ha reso disponibili i manoscritti digitalizzati, in una versione a più bassa risoluzione, in un apposito sito che permette a chiunque di ammirare gratuitamente questi capolavori.

Riferimenti bibliografici

Siti della BAV dedicati al Progetto di Digitalizzazione:

<https://www.vaticanlibrary.va/home.php?pag=digitalizzazione&ling=eng>

<https://digi.vatlib.it>

Home page del FITS

https://fits.gsfc.nasa.gov/fits_home.html

Autore



Giuseppe Di Persio - giuseppe.dipersio@inaf.it

Nasce a Roma nel 1964. Si diploma in elettronica, ma l'informatica diventa presto la sua passione. Dal 1984 lavora al CNR presso diversi istituti, per poi confluire nel 2003 all'INAF a Roma. Nel 2002 si laurea in Informatica.

È Computer Scientist presso lo Space Science Data Center dell'ASI e fa parte del team software di diversi satelliti. Si occupa di programmazione, acquisizione e analisi dati, progettazione e gestione di data center scientifici, sicurezza e reti. Negli ultimi anni anche di Long Term Data Preservation, collaborando con la Biblioteca Vaticana nel suo progetto di digitalizzazione tramite l'uso di standard astronomici.