DigItalia

Anno VI, Numero 2 - **2011**
ISSN 1972-6201

Rivista del digitale nei beni culturali

ICCU-ROMA

Analisi del formato FITS per la conservazione a lungo termine dei manoscritti. *Il caso significativo del progetto della Biblioteca Apostolica Vaticana*

Stefano Allegrezza
Università di Udine

La Biblioteca Apostolica Vaticana ha avviato un imponente progetto di digitalizzazione di oltre ottantamila preziosissimi manoscritti; per la conservazione nel tempo delle immagini acquisite ha deciso di affidarsi al formato FITS in luogo dei più consueti formati immagine (TIFF, BMP, JPG, etc.). Si tratta di una scelta che a priva vista può apparire quantomeno inconsueta, dal momento che il formato FITS è impiegato da oltre trenta nel campo astronomico ma, quasi certamente, non è mai stato utilizzato per archiviare e conservare nel tempo le immagini acquisite a seguito di digitalizzazione. Tuttavia, andando ad analizzare le specifiche del formato si scopre che possiede delle caratteristiche che ne fanno un formato pienamente compatibile con un processo di conservazione digitale. Il presente contributo, prendendo le mosse dal caso concreto del progetto di digitalizzazione della Biblioteca Apostolica Vaticana, si propone di analizzare le caratteristiche del formato FITS per cercare di comprendere se può aspirare a diventare il "formato standard" da adottare nei progetti di digitalizzazione al fine di garantire la conservazione a lungo termine dei patrimoni digitali, già acquisiti o da acquisire, da parte di biblioteche ed archivi.*

Introduzione

Il formato FITS (*Flexible Image Transport System*), sebbene abbia ormai più di trenta anni[1] e sia ampiamente utilizzato in diversi ambiti scientifici, è diventato noto negli ambienti bibliotecari solo recentemente a seguito dell'annuncio[2], fatto dal-

* Si ringrazia l'ingegnere Luciano Ammenti, responsabile dei Servizi informatici della Biblioteca Apostolica Vaticana, per le preziose informazioni fornite durante la stesura dell'articolo.

[1] La data di nascita ufficiale del formato viene fatta risalire al 28 marzo 1979, giorno in cui Donald C. Wells e Eric Greisen stipularono il primo accordo sugli aspetti tecnici del nuovo formato. Si veda l'e-mail inviata da Donald C. Wells a Eric Greisen e disponibile all'indirizzo http://fits.gsfc.nasa.gov/birthday.news.

[2] L'annuncio è stato dato il 24 marzo 2010 mediante la pubblicazione sull'Osservatore Romano della notizia "Un'iniziativa della Biblioteca Apostolica Vaticana. Manoscritti digitali", a cura di Mons. Cesare Pasini, attuale prefetto della Biblioteca Apostolica Vaticana. Il testo della notizia è stato poi riportato anche nella

la Biblioteca Apostolica Vaticana, della sua adozione come formato per la conservazione a lungo termine delle immagini digitali ottenute dalla scansione di oltre ottantamila manoscritti conservati nella Biblioteca[3]. Si tratta, com'è ovvio, di un progetto imponente che si affianca ad altri progetti di digitalizzazione in corso[4]. Tuttavia, fino ad oggi l'accento era posto più sull'obiettivo di diffusione via rete della cultura che non su quello della conservazione di tali preziosi materiali digitali. Questo a causa dei tanti problemi ancora non risolti che la conservazione dei documenti in formato digitale comporta: fragilità dei supporti informatici, obsolescenza dell'hardware e del software, obsolescenza dei formati elettronici, etc. Il formato FITS sembra promettere una soluzione ad almeno uno degli ostacoli principali che si incontrano nella conservazione digitale, quello dell'obsolescenza dei formati elettronici.
Ma di che cosa si tratta esattamente? Quali sono le caratteristiche del formato FITS che lo hanno reso così interessante da essere adottato dalla Biblioteca Apostolica Vaticana al posto di altri formati per immagini che vengono normalmente utilizzati[5]? Può essere ancora adeguato per gli scopi che la *digital preservation* si prefigge un formato che è nato oltre trenta anni fa? In questo saggio cercheremo di fornire una risposta illustrando le principali caratteristiche del formato e il suo campo di applicazione; cercheremo, inoltre, di comprendere se si tratti di un formato che può essere adottato per i progetti di digitalizzazione e soprattutto se sia compatibile con un processo di conservazione digitale.

edizione straordinaria n. 5/2010 della *Newsletter* della Biblioteca Apostolica Vaticana. La *Newsletter* è disponibile all'indirizzo http://www.vaticanlibrary.va/home.php?pag=newsletter_art_00086.

[3] La Biblioteca Apostolica Vaticana possiede una delle raccolte di testi antichi e libri rari più importanti del mondo. Essa contiene ad oggi 1.600.000 libri a stampa; 8.300 incunaboli (di cui 65 in pergamena); 80.000 manoscritti, 75.000 volumi di carte, 100.000 incisioni, 300.000 monete e medaglie; circa 20.000 oggetti di arte. Tra i pezzi più famosi della Biblioteca c'è il *Codex Vaticanus*, il manoscritto più antico e completo della Bibbia che si conosca. Dal 1985 è disponibile un catalogo informatico consultabile in linea dei volumi a stampa moderni. Il sito *web* è raggiungibile all'indirizzo http://www.vaticanlibrary.va.

[4] Per un panorama delle maggiori collezioni di manoscritti digitalizzati presenti nella rete e delle strategie che i diversi Paesi perseguono al riguardo si rimanda all'articolo di Francesca Niutta, *Manoscritti nella rete*, in «DigItalia, Rivista del digitale nei beni culturali», Anno V, Numero 2, 2010.

[5] Interessante a questo proposito è quanto affermato da Luciano Ammenti, responsabile dei Servizi Informatici della Biblioteca Apostolica Vaticana, nel corso di un'intervista riportata nell'articolo "Un archivio digitale 'spazialÈ per la Biblioteca Vaticana" pubblicato il 20 aprile 2010 nella sezione 'Tecnologia' della versione *on-line* del quotidiano La Repubblica: «Avevamo già fatto un tentativo di memorizzazione in formato digitale dei nostri manoscritti negli anni '90, con l'IBM e non abbiamo ottenuto i risultati sperati; l'IBM, infatti, non è riuscita a capire fino in fondo le nostre esigenze né in ambito paleografico né di conservazione. E così per parecchi anni abbiamo abbandonato il progetto, anche se l'esigenza rimaneva, ed era quella di trovare un formato abbastanza longevo da sopravvivere alle innovazioni tecnologiche e che operasse in ambito *open source*, cioè che fosse gratuito. Utilizzare, infatti, formati come il TIFF, il PDF o il JPG 2000 potrebbe raggiungere costi inimmaginabili, anche solo se ci chiedessero un centesimo a pagina. Abbiamo anche provato a chiedere una liberatoria scritta, in particolare per il formato TIFF, ma non ce l'hanno data: progetti di questa entità non possono essere realizzati utilizzando formati a pagamento». L'articolo è disponibile all'indirizzo http://www.repubblica.it/tecnologia/2010/ 04/20/news/biblioteca_vaticana-3489668.

Il progetto di digitalizzazione della Biblioteca Apostolica Vaticana

È utile cominciare la trattazione analizzando più in dettaglio il caso, appena citato, della Biblioteca Apostolica Vaticana. Essa ha avviato un imponente progetto di digitalizzazione di oltre ottantamila preziosissimi manoscritti, tra cui opere di immenso valore come la versione greca della Bibbia (il cosiddetto *Codice B*); il *papiro Bodmer*, la trascrizione più antica - tra il 175 e il 225 dopo Cristo, meno di un secolo e mezzo dopo la loro composizione - dei Vangeli di Luca e di Giovanni; i cosiddetti *Virgilio Romano* e *Virgilio Vaticano*, i due manoscritti tardo antichi con le opere di Virgilio; l'*Evangeliario di Lorsch*, dono di Carlo Magno, il *De arte venandi cum avibus* commissionato da Federico II, testi autografi di Petrarca, alcune edizioni della *Divina Commedia*, tra cui quella 'urbinatÈ, commissionata da Federico da Montefeltro; manoscritti in latino, greco, ebraico, samaritano, aramaico, ma anche arabo, persiano, cinese, indiano, provenzale.

Si tratta di un tesoro di immenso valore, composto da pagine variamente scritte, illustrate o annotate, da «fotografare nella più alta qualità di definizione, in modo da raccogliere il maggior numero possibile di dati e non dover ritornare in futuro a ripetere un lavoro così immenso e per non mettere di nuovo a rischio carte delicatissime sempre esposte a danneggiamenti irreversibili»[6].

Per queste ragioni la Biblioteca Vaticana ha deciso di affidarsi al formato FITS, un formato non proprietario, estremamente lineare, elaborato alcuni decenni fa dalla Nasa e usato da oltre trenta anni per la conservazione dei dati inerenti le missioni spaziali e nell'ultimo decennio in astrofisica e in medicina nucleare: «esso permette la conservazione delle immagini senza problemi futuri né di tipo tecnico né di tipo economico, in quanto aggiornato sistematicamente dalla comunità scientifica internazionale»[7]. La proposta di utilizzare il formato FITS è stata fatta dall'INAF (Istituto Nazionale di Astrofisica), che ha fornito anche la consulenza tecnica, oltre che strumenti e programmi per leggere i dati, visualizzarli ed elaborarli.

Si prevede di dover riprodurre circa quaranta milioni di pagine, corrispondenti ad una mole di dati informatici stimabile nell'ordine di ben 45 PiB[8] (45 *pebibyte*, cioè circa 45 milioni di miliardi di *byte*)[9]. Un'impresa grandiosa a profitto della cultura e in particolare della tutela e conservazione dei beni affidati alla Biblioteca

[6] Cesare Pasini, *Newletter* della Biblioteca Apostolica Vaticana, edizione straordinaria n. 5/2010, cit.

[7] *Ibidem*.

[8] Qui e nel seguito si sono utilizzati i nuovi prefissi per i multipli binari (Ki, Mi, Gi, Ti, Pi, etc.) come previsto dalla norma IEC 80000-13:2008 "*Quantities and units — Part 13: Information science and technology*".

[9] Si calcola che mediamente ogni pagina digitalizzata avrà una dimensione pari a circa 1 GB (un gigabyte). Infatti, dividendo la dimensione complessiva dei dati (pari a 45 PB) per il numero complessivo di pagine da digitalizzare (40 milioni) si ottiene: 45 PB/40 milioni di pagine = 45 x 10^{15} B/ 40 x 10^{6} pagine = 1,125 x 10^{9} B/pagina = 1,125 GB/pagina.

Apostolica, nella tradizione di un servizio culturale che la Santa Sede continua a esprimere e a sviluppare nei secoli, adeguando l'impegno e le energie alle possibilità offerte dalle nuove tecnologie.

Negli anni 2009 e 2010 è stato approntato uno studio tecnico di fattibilità, con l'apporto dei migliori consulenti interni ed esterni, anche internazionali. «Ne è scaturito un progetto di alto e innovativo valore sotto diversi profili: la realizzazione delle riprese, i formati elettronici di conservazione, la sicurezza della loro stabilità nel tempo, la manutenzione e la gestione degli archivi, e così via»[10].

La realizzazione del progetto prevede un arco temporale di un decennio, suddivisibile in *tre fasi*, la prima di tre anni, la seconda di altri tre, l'ultima di quattro anni, con possibili pause fra l'una e l'altra: se motivi di vario genere, anche eventualmente motivi di raccolta di fondi, richiedessero di procedere a tappe, dopo i primi tre anni il lavoro avrebbe già una sua consistenza e compiutezza e potrebbe quindi essere temporaneamente interrotto per riprendere gli altri tre anni e poi gli altri quattro. La prima fase prevede il coinvolgimento di circa sessanta persone, tra fotografi e tutor verificatori, nella seconda e terza fase se ne prevedono ben centoventi.

Prima di avviare definitivamente il progetto, la Biblioteca Apostolica Vaticana lo ha sottoposto a test. Dai primi di marzo 2010, infatti, ha preso avvio il cosiddetto *test bed*, cioè un "banco di prova" che ha permesso di collaudare ed esaminare in tutta la sua articolazione il grande progetto studiato e formulato, così da poterne assicurare l'adeguato funzionamento quando esso verrà intrapreso in tutta la sua ampiezza. Nella prima fase del *test bed* sono stati digitalizzati 90 manoscritti, che vanno ad aggiungersi ai 23 manoscritti già digitalizzati presenti negli archivi della Biblioteca. Il lavoro di riproduzione si è avvalso di due differenti macchine, utilizzate a seconda del tipo di materiale da riprodurre: uno scanner planetario DRS 5070 della Metis Systems, gentilmente concesso in uso gratuito dalla società produttrice, e un corpo macchina digitale Hasselblad con una risoluzione di 50 Mpixel[11]. Dopo essere state scansionate e/o fotografate, le pagine sono state salvate nel formato FITS. Oltre al server che raccoglie le immagini in formato FITS acquisite dalle due macchine, sono stati installati altri due server deputati a elaborare i dati per permettere la ricerca delle immagini sia tramite la segnatura e gli elementi descrittivi del manoscritto sia e soprattutto tramite un pattern grafico: sarà possibile utilizzare un particolare software per ricercare un determinato segno grafico o una determinata immagine (ad esempio, uno stemma nobiliare) all'interno dell'intero posseduto della Biblioteca, in maniera estremamente agile ed intelligente. Quest'ultimo strumento, veramente innovativo e certamente interessante per quanti compiranno ricerche sui manoscritti della Vaticana, nasce dalla tecnolo-

[10] Cesare Pasini, *Newletter*, cit.

[11] Ovvero, cinquanta milioni di *pixel*, pari a circa cinque volte la risoluzione tipica delle macchine fotografiche digitali oggi disponibili sul mercato *consumer*.

gia della società inglese *Autonomy Systems*, leader mondiale nel campo dell'informatica e una dei principali finanziatori del *test bed*. Questo consentirà di effettuare in maniera molto più agevole tutti quegli studi che oggi debbono essere effettuati in maniera "manuale" e, soprattutto, sarà attuabile sull'intero posseduto di 80.000 manoscritti portando a conoscenze e scoperte che oggi sono impensabili.
Nel mese di giugno 2010 si è positivamente conclusa la prima fase di test. Nel mese di ottobre dello stesso anno si è potuto dare avvio ad una seconda fase che ha permesso di digitalizzare ulteriori 100 manoscritti, portando a circa 213 il numero totale di manoscritti digitalizzati, per una mole complessiva di dati informatici pari a circa 10 TiB (dieci *tebibyte*, ovvero circa diecimila miliardi di *byte*).
Per l'avvio del progetto vero e proprio si è previsto l'utilizzo di un apposito edificio esterno alla Città del Vaticano adeguatamente attrezzato per ospitare i manoscritti, il CED, le apparecchiature necessarie per l'attività di digitalizzazione e i locali dove lavoreranno da 120 a 150 persone. A garantire la massima sicurezza degli interni e degli esterni dell'edificio e del prezioso "contenuto" si è offerta la Fondazione Enzo Hruby, che si è fatta carico di sostenere economicamente gli interventi destinati alla protezione. L'edificio sarà dotato di tutti i sistemi di sicurezza necessari, dalla sistema antitrusione e di videosorveglianza ai sistemi antincendio e antiallagamento. Per rispettare i parametri di sicurezza, le immagini digitalizzate saranno conservate in due *datastorage*, uno nella Città del Vaticano e l'altro a Castel Gandolfo. L'impresa sarà certificata ISO 27001[12].
L'avanzamento del progetto, che ha costi elevatissimi, dipende in larga misura dai finanziamenti che la Biblioteca Apostolica Vaticana riesce a reperire (cosa che, in questo momento di crisi, non è per nulla semplice). Al momento il progetto di digitalizzazione dei manoscritti non risulta ancora completamente avviato, anche se è comunque aumentato il numero dei manoscritti digitalizzati, che costituiscono un primo avvio, pur limitato nei numeri, di quel progetto. Grazie ad una iniziativa dell'Università di Heidelberg, che mira a ricostruire virtualmente l'antica biblioteca del monastero di Lorsch, è stato possibile digitalizzare 142 manoscritti Palatini provenienti da quel monastero. Analogamente, grazie ad un progetto promosso dal *Chinese National Committe for the compilation of Qing History*, si è avviata la digitalizzazione di più di 600 manoscritti e di stampati cinesi inerenti la storia cinese dal Seicento agli inizi del Novecento: «nel complesso risulteranno digitalizzati circa 1300 titoli, per un totale di 100.000 fogli, di cui circa il 30% sono manoscritti e il 70% stampati xilografici»[13].

[12] Lo standard ISO/IEC 27001:2005 "*Information technology — Security techniques — Information security management systems – Requirements*" è una norma internazionale che definisce i requisiti per impostare e gestire un Sistema di Gestione della Sicurezza delle Informazioni (SGSI o ISMS, dall'inglese *Information Security Management System*), e comprende aspetti relativi alla sicurezza logica, fisica ed organizzativa.

[13] Cesare Pasini, *Newletter* della Biblioteca Apostolica Vaticana, n. 9/2011, disponibile all'indirizzo http://www.vaticanlibrary.va/home.php?pag=newsletter_art_00128&BC=21.

Una volta digitalizzati, i manoscritti potranno rimanere per sempre al riparo in un bunker di cemento armato sotterraneo, senza finestre né aperture, a prova di bomba atomica, in condizioni di temperatura (tra i 20 e i 21 gradi) e umidità (tra 50 e 55%) controllate. Solo pochissimi vi avranno accesso per evitare di danneggiare le pagine depositarie di tanto sapere. Ma che al tempo stesso tutti potranno consultare via *web* grazie alla copia digitale: milioni e milioni di pagine virtuali che daranno al mondo scientifico e accademico possibilità di studio finora mai esplorate. Certamente l'accesso ai manoscritti originali non sarà del tutto impedito: gli studiosi che avranno la necessità di vedere l'originale potranno continuare a farlo. Ma la possibilità di accedere alla versione digitale direttamente tramite collegamenti in rete permetterà in gran parte dei casi di risparmiare al manoscritto una prolungata consultazione diretta, considerando che ogni uso, anche il solo avvicinarsi con tutte le cure possibili, può sempre rischiare di danneggiarlo. Per poter consultare in rete i primi manoscritti non bisognerà attendere dieci anni: la Biblioteca Apostolica Vaticana conta di poter inserire in rete le prime immagini già nel corso del 2012[14].

Tutto questo sarà possibile grazie alla scelta, da parte della Biblioteca Apostolica Vaticana, del formato FITS, che nel corso degli anni ha mostrato ampiamente la sua affidabilità e la sua longevità digitale. La scelta di questo formato fornisce molti spunti di riflessione e soprattutto pone tutti coloro che si occupano di digitalizzazione nel settore dei beni culturali di fronte ad un grosso interrogativo: il formato FITS andrà a rivoluzionare le future scelte in merito al formato da utilizzare nei progetti di digitalizzazione, che tradizionalmente si basano sull'utilizzo di altri formati[15] per la conservazione delle immagini ottenute dalle scansioni? Per rispondere a questa e alle domande correlate, cerchiamo innanzitutto di comprendere meglio le caratteristiche del formato.

[14] Il tutto secondo l'ispirazione originaria di Niccolò V, al secolo Tommaso Parentucelli, il grande pontefice umanista che a metà del Quattrocento fu il vero fondatore della Vaticana, poi istituita da Sisto IV nel 1475. L'intento è ben sintetizzato da Mons. Cesare Pasini: «I manoscritti sono un patrimonio dell'umanità che vogliamo rendere accessibile all'umanità e che di fronte all'umanità abbiamo la responsabilità di conservare». Si veda l'articolo pubblicato sul Corriere della Sera del 21 ottobre 2011, disponibile all'indirizzo http://www.corriere.it/cultura/libri/11_ottobre_21/ vecchi-pergamene-vaticane-bunker_741b5958-fbc2-11e0-a389-b44dd5e172d2.shtml.

[15] Si veda, ad esempio, le *Linee guida tecniche per i programmi di creazione di contenuti culturali digitali. Versione italiana 1.8*, a cura di Giuliana De Francesco, Ministero per i beni e le attività culturali, Progetto MINERVA, WP4 – Gruppo di lavoro italiano "Interoperabilità e servizi", disponibile all'indirizzo http://www.minervaeurope.org/structure/workinggroups/servprov/documents/technicalguidelinesita1_8.pdf. A pag. 35 si legge: «Le immagini a matrice di punti […] dovrebbero perciò essere memorizzate utilizzando formati quali: Tagged Image File Format (TIFF), Portable Network Graphics (PNG), oppure, con attenzione alle considerazioni che seguono, anche Graphical Interchange Format (GIF), JPEG 2000 o JPEG Still Picture Interchange File Format (JPEG/SPIFF)». Come si vede, non vi è alcun riferimento al formato FITS.

Il formato FITS

FITS è il formato per i dati scientifici più utilizzato nel campo dell'astronomia ed è diffuso anche nel campo dell'astrofisica. Il formato era inizialmente utilizzato per l'interscambio di immagini tra piattaforme hardware e applicazioni software che non condividevano un formato di *file* comune, ma oggi è diventato il formato standard per l'archiviazione e la conservazione delle immagini astronomiche di molte organizzazioni scientifiche: si pensi, ad esempio, alle immagini ottenute da radiotelescopi ed altre apparecchiature astronomiche e a quelle restituite dai satelliti orbitali, dalle navicelle spaziali e dalle sonde planetarie nel corso delle loro missioni (ad esempio, la NASA lo utilizza per conservare le immagini delle missioni spaziali). Da qualche anno comincia ad essere adoperato anche in astrofisica ed in medicina nucleare.

È bene far notare subito che FITS non è solamente *uno* dei tanti formati per immagini di tipo *raster*[16], (come il TIFF, il BMP, il JPG, il PCX, il TGA, etc.) ma è un formato molto più generale in grado di memorizzare anche altri tipi di dato: testo codificato in ASCII, dati scientifici organizzati in matrici bidimensionali e multidimensionali[17], etc. A differenza di altri formati per immagini, FITS è stato progettato specificamente per dati di tipo scientifico e prevede, quindi, la possibilità di utilizzare metadati per descrivere compiutamente l'oggetto rappresentato ed il contesto di produzione.

Evoluzione del formato FITS

Il formato FITS è diffusissimo come formato per le immagini nei campi dell'astronomia e dell'astrofisica, ma il suo impiego in progetti di digitalizzazione di manoscritti è stato, fino ad oggi, un evento piuttosto raro (e certamente non era stato previsto da coloro che lo avevano sviluppato).

La data di nascita del formato viene fatta risalire al 1979[18] ma solo nel 1981 con la pubblicazione sulla rivista *Astronomy and Astrophysics Supplement* dell'articolo "*FITS: a flexible image transport system*" vennero ufficialmente presentate le sue specifiche. Gli autori dell'articolo, Eric W. Greisen, Don C. Wells e Ron H. Harten, con l'aggiunta di un quarto autore, P. Grosbol, descrissero il formato anche in altri tre articoli pubblicati sulla stessa rivista tra il 1981 e il 1988.

[16] I formati per le immagini possono essere classificati sulla base della tecnologia utilizzata per rappresentare l'immagine stessa. Si hanno in questo modo tre principali categorie: *raster* (o *bitmap*), *vector* (o vettoriali) e *ibride* (*hybrid*). I formati di tipo *raster* organizzano le informazioni in una griglia rettangolare (*raster*) di *pixel* (= *picture element*), ciascuno dei quali è rappresentato con un certo numero di *bit* a seconda dello profondità di colore che si intende utilizzare (bianco e nero, scala di grigi, colori a 16 *bit*, colori a 24 *bit*, etc.).

[17] Come sarà chiaro più avanti, i *file* FITS non necessariamente contengono immagini: infatti possono contenere anche altre tipologie di dati (in generale di tipo scientifico).

[18] Si veda la nota 1.

Complessivamente questi articoli sono conosciuti come "*Four FITS papers*"[19].
La motivazione principale all'introduzione del formato derivava dal fatto che all'epoca la maggior parte delle organizzazioni scientifiche operanti nel campo dell'astronomia utilizzava, per memorizzare i dati scientifici, formati di *file* di tipo proprietario e non supportati dai software in uso presso le altre organizzazioni. Era sentita, pertanto, l'esigenza di superare la situazione di insufficiente interoperabilità allora diffusa[20] grazie ad un formato comune che consentisse l'interscambio dei dati. Proprio per rispondere a questa esigenza, gli autori dei "*Four FITS papers*" pubblicarono, nel 1981, le prime specifiche del formato FITS.
In seguito la gestione del formato FITS passò ad un organismo ufficiale, il *NASA/OSSA Office of Standards and Technologies* (NOST) presso il *Goddard Space Flight Center*, che, basandosi degli articoli sopra citati, formalizzò il formato FITS pubblicandone le specifiche in una serie di documenti sotto forma di *draft*. Questi documenti cominciarono a circolare negli ambienti scientifici e, in poco tempo, il formato FITS, divenne uno dei formati di interscambio standard *de facto*. Tuttavia la prima versione ufficiale, denominata *NOST 100-1.0*, venne formalizzata solo nel giugno del 1993 dallo *IAU FITS Working Group* (IAUFWG)[21]. La seconda versione ufficiale, denominata *NOST 100-2.0*, risale al mese di marzo del 1999. Questa versione venne pubblicata sulla rivista scientifica *Astronomy and Astrophysics* nel mese di settembre del 2001. La versione più recente, la numero 3.0, è stata ufficialmente approvata dallo IAUFWG nel mese di luglio del 2008[22].

[19] Gli articoli in questione sono precisamente: D.C. Wells, E.W. Greisen, e R.H. Harten, *FITS: a flexible image transport system*, in Astronomy and Astrophysics Supplement Series, vol. 44, p. 363-370, giugno 1981; E.W. Greisen, R.H. Harten, *An extension of FITS for small arrays of data*, in Astronomy and Astrophysics Supplement Series, vol. 44, p. 371-374, 1981; P. Grosbol, R.H. Harten, E.W. Greisen, D.C. Wells, *Generalized extensions and blocking factors for FITS*, in Astronomy and Astrophysics Supplement Series, vol. 73, p. 359-364, 1988; R.H. Harten, P. Grosbol, E.W. Greisen, D.C. Wells, *The FITS tables extension*, in Astronomy and Astrophysics Supplement Series, vol. 73, p. 365-372, 1988.

[20] Cfr. R. J. Hanisch, A. Farris, E. W. Greisen, W. D. Pence, B. M. Schlesinger, P. J. Teuben, R. W. Thompson and A. Warnock III, *Definition of the Flexible Image Transport System (FITS)*, Astronomy & Astrophysics, Volume 376, Number 1, September 2001.

[21] Lo *IAU FITS Working Group* (IAUFWG) è l'autorità di controllo internazionale per il formato FITS costituita in seno all'*International Astronomical Union* (IAU). È stato costituito nel 1988 mediante una risoluzione formale della XX Assemblea Generale dello IAU a Baltimora allo scopo di per mantenere gli standard FITS esistenti e di approvare le future estensioni del formato. Il sito *web* dello IAUFWG è raggiungibile all'indirizzo http://fits.gsfc.nasa.gov/iaufwg

[22] La specifica della versione 3.0 è disponibile sia come documento ufficiale (*Definition of the Flexible Image Transport System (FITS). The FITS Standard. Version 3.0*), approvato il 10 luglio 2008 dallo IAUFWG, sia come articolo pubblicato su rivista (W. D. Pence1, L. Chiappetti, C. G. Page, R. A. Shaw, E. Stobie, *Definition of the Flexible Image Transport System (FITS), version 3.0*, in Astronomy and Astrophysics, n. 524, Dicembre 2010). Entrambi sono disponibili alla pagina http://fits.gsfc.nasa.gov/fits_standard.html.

La **Tabella 1** riporta le varie versioni del formato che si sono succedute a partire dalla sua nascita. Come si può osservare, si tratta di un formato piuttosto stabile, dal momento che nel corso di circa trenta anni si sono avvicendate solamente tre versioni principali[23].

Le specifiche del formato FITS sono descritte in un documento denominato "FITS Standard"[24]. L'ultima versione, quella che descrive il formato 3.0, è costituita da un testo di circa una cinquantina di pagine suddiviso in sezioni che descrivono compiutamente ogni aspetto del formato: l'organizzazione del *file*, le intestazioni (*header*), le parole chiave (*keyword*), la rappresentazione dei dati, le estensioni standard (*Image extensions*, *ASCII Table Extensions*, *Binary Table Extensions*)[25], il sistema di coordinate mondiali, etc. Attualmente la gestione del formato è a cura dello IAUFWG[26].

Versione del formato	Data di rilascio	Note
-	1981	Nascita del formato
NOST 100-0.1	Dicembre 1990	Primo *draft* dello Standard
NOST 100-0.2	Giugno 1991	Secondo *draft* dello Standard
NOST 100-0.3	Dicembre 1991	Terzo *draft* dello Standard
NOST 100-1.0	Giugno 1993	Standard NOST (prima versione)
NOST 100-1.1	Settembre 1995	Standard NOST
NOST 100-1.2	Aprile 1998	Alcune modifiche di minore entità rispetto alla precedente versione
NOST 100-2.0	Marzo 1999	Standard NOST (seconda versione, superata ma ancora supportata)
IAUFWG 2.1	Aprile 2005	Standard IAUFWG
IAUFWG 2.1b	Dicembre 2005	Standard IAUFWG (aggiunge il supporto ai numeri interi a 64 bit)
IAUFWG 3.0	Luglio 2008	Standard IAUFWG (terza versione)

Tabella 1. Le versioni del formato FITS

[23] Al confronto, altri formati elettronici sono molto meno stabili. Si pensi che il formato PDF ha avuto dal momento della sua nascita, avvenuta nel 1992, ad oggi ben otto versioni. Lo stesso si può dire di altri formati piuttosto noti come il formato DOC (Microsoft Word) o il formato XLS (Microsoft Excel), oggi entrambi soppiantati dai loro successori, i formati DOCX e XLSX (i formati di *default* a partire dalla versione 2007 di Microsoft Word ed Excel) e, quindi, destinati all'obsolescenza.

[24] Le varie versioni del *FITS Standard* sono disponibili sul sito del *FITS Support Office*, all'indirizzo http://fits.gsfc.nasa.gov/fits_standard.html.

[25] Si veda più avanti al paragrafo 5.

[26] La gestione del formato FITS rientra anche tra i principali compiti del *Working Group on Astronomical Software* (WGAS) della *American Astronomical Society* (AAS).

Analisi tecnica del formato FITS[27]

Nonostante la lettera "I" presente nel nome del formato (*Flexible "I"mage Transport System*) sia l'iniziale del termine "Image" (immagine), in realtà il formato FITS non consente solo l'interscambio e la conservazione di immagini ma è di utilizzo molto più generale[28]. Infatti, nelle specifiche del formato non c'è alcun vincolo che ne limiti l'utilizzo alle sole immagini per cui può ben essere considerato un formato di memorizzazione dei dati di utilizzo generico.

Uno dei punti di forza del formato va ricercato senz'altro nella sua semplicità: la struttura di un *file* codificato secondo il formato FITS è, infatti, estremamente lineare. Esso è composto da una sequenza di uno o più blocchi (si veda la **Figura 1**), ciascuno dei quali è denominata HDU (*Header and Data Unit*): il primo è denominato "*Primary HDU*" ed i successivi sono denominati "*Extension HDU*". Ciascun blocco HDU a sua volta è composto da due sezioni: una intestazione (detta *Header Unit* od anche, semplicemente, *Header*) ed una sezione, opzionale, contenente dati (detta *Data Unit* od anche *Data Array*)[29]; nel caso della *Primary HDU* l'intestazione prende il nome di *Primary Header Unit* e la sezione dati prende il nome di *Primary Data Unit* (od anche "*Primary Data Array*"); invece, nel caso delle *Extension HDU* le due sezioni prendono il nome di *Extension Header Unit* e *Extension Data Unit* (o *Extension Data Array*), rispettivamente.

[27] Per un approfondimento si rimanda alla documentazione presente sul sito ufficiale della NASA dedicato al formato FITS (*FITS Support Office at NASA/GSFC*), all'indirizzo http://fits.gsfc.nasa.gov. In particolare si segnalano i documenti *A Brief Introduction to FITS* disponibile all'indirizzo http://fits.gsfc.nasa.gov/fits_overview.html e *A primer on the FITS Data Format*, disponibile all'indirizzo http://fits.gsfc.nasa.gov/fits_primer.html. Si veda anche la *Encyclopedia of Graphics File Formats*, 2nd Edition, di James D. Murray, William Van Ryper, ISBN: 1-56592-161-5. La versione *on-line*, distribuita da O'Reilly secondo la licenza Creative Common/Attribution, è disponibile all'indirizzo http://www.fileformat.info/mirror/egff/index.htm; la pagina *FITS File Format Summary* dedicata al formato FITS è raggiungibile all'indirizzo http://www.fileformat.info/format/fits/egff.htm. Si veda anche: *Overview of the FITS Data Format* sul sito del *High Energy Astrophysics Science Archive Research Center* (HEASARC) della *National Aeronautics and Space Administration* (NASA), disponibile all'indirizzo http://heasarc.gsfc.nasa.gov/docs/heasarc/ fits_overview.html; NASA/GSFC Astrophysics Data Facility; *A User's Guide for the Flexible Image Transport System (FITS). Version 4.0*, 1997, disponibile all'indirizzo http://fits.gsfc.nasa.gov/users_guide/usersguide.pdf. Si veda, infine, l'archivio dei materiali messo a disposizione dal *National Radio Astronomy Observatory* (NRAO) all'indirizzo http://www.cv.nrao.edu/fits. Uno studio sul formato è stato condotto anche nell'ambito del progetto InterPARES2 all'interno del *Case Study 26: Microvariability & Oscillations of Stars* (MOST) *Satellite Mission: Preservation of Space Telescope Data*, a cura di Peter Gagné, Maggio 2006, che ha preso in esame la conservazione dei dati scientifici nel campo dell'indagine telescopica con particolare riguardo all'aspetto dell'autenticità.

[28] Ovvero, appartiene alla categoria dei formati di utilizzo generale (*general data file format*).

[29] Normalmente una HDU contiene una *Header Unit* seguita da una *Data Unit*, ma le specifiche del formato prevedono anche il caso in cui una HDU contenga solamente una *Header Unit*.

Il più semplice *file* in formato FITS è costituito da una sola *Primary HDU* e può contenere matrici di dati di dimensione da 1 a 999. Nel caso di dimensione "2" la matrice viene detta bidimensionale ed i dati costituiscono tipicamente la rappresentazione digitale di una immagine in formato *raster* (è questo il caso del progetto di digitalizzazione della Biblioteca Apostolica Vaticana).
La struttura del formato è flessibile e consente di aggiungere, di seguito alla *Primary HDU*, un numero qualsiasi di *Extension HDU* a seconda della quantità di informazioni che è necessario memorizzare. Attualmente le specifiche del formato prevedono tre tipi standard di *Extension HDU*:

- *Image Extension HDU*, utilizzata per memorizzare matrici di *pixel* di dimensione compresa tra 0 e 999 (esattamente la stessa tipologia di dato memorizzata nella *Primary HDU*). In questo caso la *Header Unit* inizia con la stringa XTENSION = 'IMAGÈ).
- *ASCII Table Extension HDU*, utilizzata per memorizzare informazioni sotto forma di tabella utilizzando dati codificati in ASCII[30]. In questo caso la *Header Unit* inizia con la stringa XTENSION = 'TABLÈ. Le tabelle ASCII sono generalmente meno efficienti rispetto alle tabelle binarie, ma possono essere costruite in maniera da essere facilmente leggibili dall'uomo (*human readable*) e sono in grado di memorizzare informazioni numeriche con dimensione e precisione arbitrarie (ad esempio, numeri reali a 16 *byte*).
- *Binary Table Extension HDU*, utilizzata per memorizzare informazioni sotto forma di tabella utilizzando questa volta una rappresentazione binaria. In questo caso la *Header Unit* inizia con la stringa XTENSION = 'BINTABLÈ. Ogni cella della tabella può essere una matrice, ma la dimensione di tale matrice deve essere costante all'interno di una colonna. Grazie a questo tipo di estensione è possibile utilizzare il formato FITS per memorizzare, ad esempio, interi database relazionali.

In aggiunta alle strutture di cui sopra, c'è un altra tipologia di HDU denominata '*Random Group*' che è utilizzata quasi esclusivamente per applicazioni di interferometria radio e il cui utilizzo è sconsigliato per altre tipologie di applicazione all'infuori di questa.
Grazie alle *Extension HDU* è possibile utilizzare un *file* FITS per memorizzare oggetti digitali di vario tipo (non solo immagini, ma anche spettri, cubi di dati, dati

[30] L'ASCII (*American Standard Code for Information Exchange*) è una codifica dei caratteri. L'ASCII *standard* (detto anche US-ASCII) utilizza 7 *bit* per ogni carattere, riuscendo così a rappresentare 2^7 = 128 caratteri diversi: i numeri da 0 a 9, l'alfabeto romano maiuscolo e minuscolo, i segni di punteggiatura e una serie di caratteri speciali come, ad esempio, '@'. L'ASCII standard è il minimo comun denominatore di ogni comunicazione testuale in forma digitale. Un *file* codificato in ASCII standard può essere "compreso" da qualsiasi computer, indipendentemente dal programma utilizzato. L'ASCII *esteso* utilizza invece 8 *bit* per codificare ciascun carattere riuscendo così a rappresentare 2^8 = 256 caratteri diversi.

strutturati come database multi-tabellari, etc.); inoltre, è possibile memorizzare in un unico *file* FITS più oggetti digitali (anche di categorie diverse) una per ciascuna *Extension*. Ad esempio, se si volesse memorizzare tre immagini all'interno di in un unico *file* FITS, si potrebbe utilizzare la *Primary HDU* per memorizzare la prima immagine e due *Extension HDU* per memorizzare le altre due immagini. Allo stesso modo, è possibile memorizzare nello stesso *file* FITS una radiografia ed un esame ecografico, utilizzando per ciascuna tipologia di risorsa digitale una diversa *Extension*.

Per ragioni storiche[31], la dimensione di ogni *Header Unit* ed ogni *Data Unit* deve essere un multiplo esatto di 2.880 *byte* (corrispondente a 36 stringhe alfanumeriche della lunghezza di 80 *byte* ciascuna). Se necessario, ogni spazio inutilizzato alla fine di una *Header Unit* o di una *Data Unit* deve essere completato con caratteri di riempimento fino a raggiungere tale dimensione: nel caso della *Header Unit* va utilizzato il carattere ASCII "SP" (32 decimale o 20 esadecimale[32])[33], mentre nel caso della *Data Unit* si utilizza il carattere ASCII "NUL" (0 decimale o 00 esadecimale)[34].

Lo standard FITS ha avuto origine in un'epoca in cui i *file* erano archiviati e scambiati tramite nastri magnetici, pertanto non prescrive alcuna regola per la denominazione dei *file* su disco. Di conseguenza un *file* in formato FITS è valido anche se il suo nome non presenta alcuna estensione. All'interno delle comunità scientifiche nelle quali è utilizzato il formato si è da tempo stabilita la convenzione pratica secondo la quale è possibile capire se un *file* è in formato FITS in base al contesto in cui si trova; in questi casi l'estensione del *file* viene comunemente utilizzata per altri scopi, ad esempio per specificare il contenuto del *file* anziché il suo formato. Tuttavia, in assenza di informazioni di contesto, è bene utilizzare l'estensione *.fits* (la meno ambigua), oppure altre estensioni come *.fit* e *.fts*. Il tipo MIME del formato è image/fits oppure application/fits.

[31] La dimensione di 2.880 *byte* è legata al fatto che il formato è nato per lo scambio dei dati mediante i supporti di memorizzazione all'epoca più diffusi, i nastri magnetici a 9 tracce, i quali prevedevano una strutturazione dei dati in blocchi logici. Si veda la pagina *FITS Files and Physical Media* all'indirizzo http://archive.stsci.edu/fits/ users_guide/node49.html.

[32] La codifica esadecimale (spesso abbreviata come "esa" o "hex") si basa su un sistema numerico posizionale in base 16 ed utilizza 16 simboli invece dei 10 del sistema numerico decimale tradizionale. In genere si utilizzano i simboli da 0 a 9 per le prime dieci cifre, e poi le lettere da A a F per le successive sei cifre, per un totale di 16 simboli.

[33] Si tratta del carattere "spazio". Si noti che in una *Header Unit* non possono essere utilizzati i caratteri ASCII cosiddetti "non stampabili" o "caratteri di controllo", ovvero i caratteri con valori decimali da 0 a 32 e il 127. Tra di essi vi sono i caratteri NUL (0 decimale o 00 esadecimale), HT (*horizontal tab*) (9 decimale o 09 esadecimale), LF (*line feed*) (10 decimale o 0A esadecimale), VT (*vertical tab*) (11 decimale o 0A esadecimale), CR (*carriage return*) (13 decimale o 0D esadecimale) e il carattere DEL (127 decimale o 7F esadecimale). Si veda la **Tabella 2**.

[34] Il carattere ASCII "NUL" è composto da una sequenza di otto *bit* "zero".

Struttura della Header Unit

Ogni *Header Unit* è composta da uno o più *record*, denominati *keyword record* (od anche, con terminologia un po' obsoleta, *card image*)[35], ciascuno dei quali ha lunghezza obbligatoria di 80 caratteri ed è codificato utilizzando il set ristretto di caratteri compreso tra il carattere ASCII "SP" e il carattere ASCII "~" (126 decimale o 7E esadecimale)[36]; non sono consentiti i cosiddetti "caratteri ASCII non stampabili" come le tabulazioni, il ritorno a capo o il *line-feed* (si veda la **Tabella 2**).

I *keyword record* sono molto importanti perché forniscono informazioni sui dati memorizzati nella successiva *Data Unit* (ad esempio, la dimensione, il formato, l'autore, etc.). In sostanza si tratta di metadati che possono essere inclusi all'interno dello stesso *file* (ed è per questo che il formato gode dell'importante caratteristica di essere auto-documentato)[37]. La sintassi di un *keyword record* è la seguente:

ASCII control			ASCII text								
dec	hex	char	dec	hex	char	dec	hex	char	dec	hex	char
0	00	NUL	32	20	SP	64	40	@	96	60	`
1	01	SOH	33	21	!	65	41	A	97	61	a
2	02	STX	34	22	"	66	42	B	98	62	b
3	03	ETX	35	23	#	67	43	C	99	63	c
4	04	EOT	36	24	$	68	44	D	100	64	d
5	05	ENQ	37	25	%	69	45	E	101	65	e
6	06	ACK	38	26	&	70	46	F	102	66	f
7	07	BEL	39	27	'	71	47	G	103	67	g
8	08	BS	40	28	(	72	48	H	104	68	h
9	09	HT	41	29	)	73	49	I	105	69	i
10	0A	LF	42	2A	*	74	4A	J	106	6A	j
11	0B	VT	43	2B	+	75	4B	K	107	6B	k
12	0C	FF	44	2C	,	76	4C	L	108	6C	l
13	0D	CR	45	2D	-	77	4D	M	109	6D	m
14	0E	SO	46	2E	.	78	4E	N	110	6E	n
15	0F	SI	47	2F	/	79	4F	O	111	6F	o
16	10	DLE	48	30	0	80	50	P	112	70	p
17	11	DC1	49	31	1	81	51	Q	113	71	q
18	12	DC2	50	32	2	82	52	R	114	72	r
19	13	DC3	51	33	3	83	53	S	115	73	s
20	14	DC4	52	34	4	84	54	T	116	74	t
21	15	NAK	53	35	5	85	55	U	117	75	u
22	16	SYN	54	36	6	86	56	V	118	76	v
23	17	ETB	55	37	7	87	57	W	119	77	w
24	18	CAN	56	38	8	88	58	X	120	78	x
25	19	EM	57	39	9	89	59	Y	121	79	y
26	1A	SUB	58	3A	:	90	5A	Z	122	7A	z
27	1B	ESC	59	3B	;	91	5B	[	123	7B	{
28	1C	FS	60	3C	<	92	5C	\	124	7C	|
29	1D	GS	61	3D	=	93	5D	]	125	7D	}
30	1E	RS	62	3E	>	94	5E	^	126	7E	~
31	1F	US	63	3F	?	95	5F	_	127	7F	DEL

KEYWORD = value / comment

dove:

– *keyword* è una stringa che specifica il formato e il significato dei dati memorizzati nel *keyword record*. Essa può essere lunga da uno ad otto caratteri ed occupa sempre le posizioni da 1 fino ad 8 del *keyword record*.

Può contenere solamente le lettere maiuscole dalla A alla Z, le cifre da 0 a 9, il trattino (-) ed il trattino basso

Tabella 2. Il set di caratteri ASCII (fonte: W. D. Pence, L. Chiappetti, C. G. Page, R. A. Shaw, and E. Stobie, Definition of the Flexible Image Transport System (FITS), version 3.0, in Astronomy and Astrophysics, numero 524, Anno 42 (2010) - DOI: 10.1051/0004-6361/201015362).

[35] Con *keyword* si intendeva nel passato una stringa di caratteri ASCII solitamente della lunghezza di 80 caratteri. Il nome si ricollega alle schede perforate (*punched card*) utilizzate fino alla fine degli anni settanta. Le schede perforate sono state prodotte in diversi formati, ma quello più diffuso è stato il formato IBM che prevedeva la presenza di 80 colonne con 12 posizioni di perforatura ciascuna e consentiva di rappresentare esattamente 80 caratteri. Il formato UNIVAC consentiva, invece, di memorizzare 90 caratteri. Molti formati di dati, come il formato FITS, usano ancora le *keyword* come blocchi di costruzione di base anche se le schede perforate sono obsolete da diversi decenni.

[36] Si tratta fondamentalmente dei caratteri dell'alfabeto anglosassone; non è possibile utilizzare, ad esempio, i caratteri accentati presenti nell'alfabeto italiano.

[37] Si veda il paragrafo 7.

(_)[38] ed è giustificata a sinistra. Se una *keyword* è lunga meno di otto caratteri viene completata aggiungendo degli spazi fino a raggiungere tale lunghezza. Solitamente la *keyword* è seguita dal segno di "uguale" (=) e dal carattere ASCII "SP" ("spazio") collocati nelle colonne 9 e 10 del *record*, seguito poi dal relativo valore (*value*). Alcune *keyword* (ad esempio, COMMENT e HISTORY) non sono seguite dal segno di uguale e in questo caso le colonne del record che vanno dalla posizione 9 ad 80 possono contenere qualsiasi stringa di testo ASCII.

– *value* è il valore che viene assegnato alla *keyword*, ed è la rappresentazione nella codifica ASCII del dato numerico o alfanumerico associato alla *keyword*; il *value* può essere un numero intero, un numero reale in virgola mobile, un numero complesso, una stringa di caratteri (racchiusa tra virgolette singole) o un valore booleano (la lettera T o F);
– *comment* è un campo di commento opzionale che può essere inserito di seguito ad ogni *value* nella *keyword* allo scopo di fornire informazioni e può contenere ogni carattere ASCII standard. È separato da *value* mediante il carattere "slash" ("/") o mediante uno spazio ed uno "slash" (" /") (è raccomandata la seconda sintassi).

Sono molte le *keyword* che possono essere utilizzate all'interno di una *Header Unit*[39], ma di queste solamente cinque sono obbligatorie e devono essere sempre presenti: si tratta delle *keyword* SIMPLE, BITPIX, NAXIS, NAXIS*n* e END. Inoltre, nel caso in cui siano previste *Extension Unit* deve essere presente anche la *keyword* EXTEND.
La *keyword* SIMPLE deve essere sempre la prima di ogni *Header Unit*. Il suo *value* indica il livello di conformità alle specifiche del formato FITS descritte nel "FITS Standard": se il *file* è pienamente conforme ad esso, il valore è "T" ("True"); se, invece, il *file* differisce in qualche maniera dai requisiti specificati il valore è "F" ("False")[40].
La *keyword* BITPIX consente di specificare il numero di *bit* utilizzati per rappresentare il valore di ogni dato. Il relativo *value* deve essere un numero intero che, nel caso di dati riferibili ad immagini, rappresenta il numero di *bit* utilizzati per rappresentare ciascun *pixel* dell'immagine (ovvero, la cosiddetta "profondità di colore"). I valori possibili sono riepilogati nella **Tabella 3**.

[38] Non possono essere utilizzati caratteri minuscoli, segni di punteggiatura o codici di controllo. Se in una *keyword* non è presente una keyword (ovvero, le colonne da 1 ad 8 sono occupate da spazi), la *keyword* è considerata alla stregua di un commento.
[39] È improbabile che qualsiasi lettore FITS li capiscono tutti (le *keyword* non riconosciute sono trattate come commenti dei lettori FITS. For a description of all other valid FITS header keywords, refer to the FITS specification.
[40] I due valori "T" e "F" stanno, rispettivamente, per "True" (Vero) e "False" (Falso).

Valore di BITPIX	Tipo di dato
8	Carattere o numero intero senza segno
16	Numero intero con segno a 16 bit
32	Numero intero con segno a 32 bit
64	Numero intero con segno a 64 bit
-32	Numero reale in virgola mobile a 32 bit a precisione singola
-64	Numero reale in virgola mobile a 64 bit a precisione doppia

Tabella 3. I possibili valori che può assumere il campo value nel caso del keyword record BITPIX.

La *keyword* NAXIS specifica un valore intero, compreso tra "0" e "999", che indica il numero di assi nella matrice dei dati. Ad esempio, le consuete immagini *raster* hanno un valore NAXIS pari a "2". Il valore "0" significa alla *Header Unit* non fa seguito alcuna *Data Unit* (anche se può essere comunque presente una *Extension Unit*).
La *keyword* NAXISn indica la lunghezza di ogni asse in unità BITPIX. Ad esempio, nel caso di una immagine di dimensioni 300 x 400 è necessario specificare due *keyword*, NAXIS1 con valore pari a "300" e NAXIS2 con valore pari a "400". Se il valore della *keyword* NAXIS è "0" non è presente, ovviamente, alcuna *keyword* NAXISn. Il campo *value* di questa *keyword* deve contenere un numero intero non negativo, che rappresenta il numero di posizioni lungo l'asse n di una matrice di dati. Per ogni valore n = 1, ..., NAXIS della *keyword* NAXIS deve essere presente la corrispondente *keyword* NAXISn. Se si assegna il valore "0" a ciascuna delle *keyword* NAXIS*n*, alla *Header Unit* non fa seguito alcuna *Data Unit*.
La *keyword* EXTEND deve essere specificata solo se nel *file* FITS ci sono delle *Extension HDU*.
L'ultima *keyword* presente in ogni *Header Unit* è sempre END, che è una *keyword* speciale non presentando né il campo *value* né il campo *comment*. La *keyword* è completata con caratteri ASCII "SP" dalla colonna 9 alla colonna 80.

Ciascuna *keyword* deve avere una lunghezza di 80 caratteri: nel caso in cui la sua lunghezza fosse inferiore deve essere completata con tanti caratteri ASCII "SP" quanti ne occorrono per raggiungere gli 80 caratteri. Inoltre, come già detto, ciascuna *Unit* deve avere la dimensione di 2.880 *byte* o di un suo multiplo; per questo motivo i *keyword record* sono raggruppati in gruppi da 36, in modo da formare blocchi di 2.880 caratteri (36 x 80 = 2.880, si veda la **Figura 2**).
Di conseguenza, anche la *Header Unit* è completata, se necessario, con tanti record vuoti, ovvero contenenti il solo carattere ASCII "SP", quanti ne occorrono per far si che essa raggiunga la dimensione di 2.880 *byte* o di un suo multiplo.

Oltre a quelle appena viste, vi sono altre *keyword* standard che possono essere utilizzate; tra queste si segnalano DATE e ORIGIN, che appartengono alla categoria delle *general descriptive keyword*. La *keyword* DATE è utilizzata per specificare la data ed eventualmente l'ora in cui è stata creata l'HDU. Il suo *value* è formato da una stringa di caratteri nella forma YYYY-MM-DD se si vuole indicare solo la data, oppure nella forma YYYY-MM-DDThh: mm:ss[.sss. . .][41] se si vuole indicare anche l'ora (nel qual caso il carattere "T" separa la data dall'ora).
Ad esempio, il seguente *keyword record*:

DATE = '2012-12-13T18:28:58' / Data ed ora di produzione del file

indica che il *file* FITS è stato prodotto il 13 dicembre 2012 alle ore 18, 28 minuti e 58 secondi.

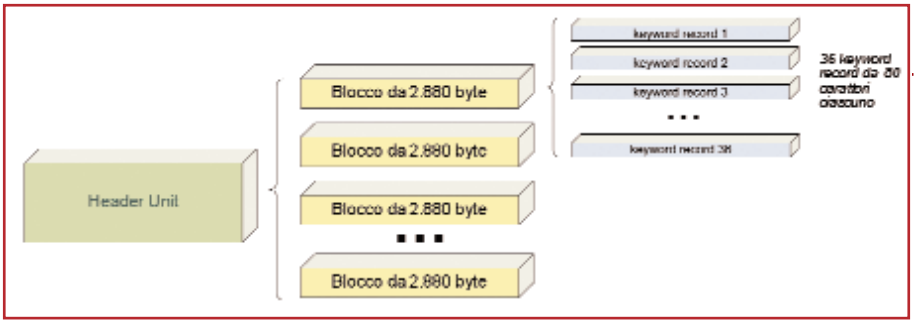

Figura 2. La struttura della Header Unit

La *keyword* ORIGIN è utilizzata per specificare l'organizzazione o l'ente responsabile della creazione del *file* FITS.
Lo standard prevede anche *biblioghaphic keyword*: AUTHOR e REFERENC. La *keyword* AUTHOR è utilizzata per specificare colui che ha compilato le informazioni presenti nella *Header Unit*. L'uso di questa *keyword* è appropriato, ad esempio, quando i dati provengono da uno studio pubblicato su una rivista; in tal caso è utilizzata per citarne il riferimento[42].

[41] Dove "YYYY" è il numero a quattro cifre dell'anno solare, "MM" il numero a due cifre del mese (con "01" equivalente a gennaio e "12" equivalente a dicembre), "DD" il numero a due cifre del giorno del mese, "hh" deve essere il numero a due cifre dell'ora del giorno, "mm" il numero a due cifre dei minuti e "ss[.sss. . .]" il numero dei secondi (due cifre seguite da una frazione opzionale). La parte decimale del campo dei secondi è opzionale e può essere arbitrariamente lunga, purché sia coerente con le regole per i formati previste dallo standard. Il *value* della *keyword* DATA deve essere espresso in UTC (*Coordinated Universal Time*).

[42] Lo standard raccomanda che vengano inclusi nel *value*, se disponibili, sia l'identificativo bibliografico a 19 cifre utilizzato nei database bibliografici del *Astrophysics Data System* (http://adswww.harvard.edu) (ad esempio: "1994A&AS..103..135A") nel caso si tratti di dati

Infine, vi sono alcune *commentary keyword* da utilizzare per fornire informazioni di commento riguardo il contenuto o la storia del *file* FITS: la *keyword* COMMENT, utilizzata per fornire commenti di tipo generico, e la *keyword* HISTORY, utilizzata solitamente per descrivere le varie elaborazioni che i dati hanno subito.
È anche possibile utilizzare un *keyword record* in cui la *keyword* è vuota (in questo caso il record inizia con il carattere "/") ed è presente solo la parte di commento: questo tipo particolare di *keyword* "vuota" è spesso utilizzato per separare gruppi di *keyword* correlate e conferire maggiore chiarezza alla sequenza di *keyword record*.
Lo standard prevede pure l'utilizzo di una serie di *keyword* destinate a fornire informazioni riguardanti il processo di osservazione astronomica quali, ad esempio: la *keyword* DATE-OBS, utilizzata per fornire informazioni riguardo la data dell'osservazione (il formato è analogo a quello della *keyword* DATE); le *keyword* TELESCOP e INSTRUME, utilizzate per fornire informazioni riguardo, rispettivamente, il telescopio e lo strumento impiegati per acquisire i dati associate con la *Header Unit*; la *keyword* OBSERVER, utilizzata per identificare il soggetto che ha acquisito i dati; la *keyword* OBJECT, utilizzata per fornire un nome dell'oggetto osservato. Si tenga presente che queste *keyword* possono essere utilizzate non solo per le osservazioni astronomiche, ma anche, ad esempio, per le acquisizioni di immagini digitali nei processi di digitalizzazione.
Oltre alle *keyword* standard, le specifiche del formato FITS prevedono l'utilizzo anche di altre *keyword* definite dall'utente; ovviamente il loro nome deve essere diverso da quello delle *keyword* previste dallo standard.

Si noti che lo standard FITS stabilisce regole molto ferree riguardo l'organizzazione dei dati.
Ad esempio:

– tutti i dati presenti in un *keyword record* devono essere obbligatoriamente collocati in specifiche colonne: l'identificativo della *keyword* (SIMPLE, BITPIX, NAXIS, etc.) occupa sempre le colonne da uno ad otto; il segno di "uguale" e il carattere di spazio che lo segue occupano sempre le posizioni nove e dieci e, se è presente, deve necessariamente seguire un valore; un valore booleano (ad esempio, "T" e "F") occupa sempre la posizione 30; i numeri interi e i valori in virgola mobile sono sempre collocati nelle colonne dalla 11 alla 30 e sono giustificati a destra inserendo, se necessario, dei caratteri ASCII "SP"; i numeri complessi interi e i numeri complessi in virgola mobile[43] sono posizionati

astronomici, sia il *Digital Object Identifier* (<http://doi.org>) (ad esempio: "doi:10.1006/jmbi.1998.2354").

[43] Si ricorda che i numeri complessi sono numeri nella forma $a + bi$, dove a e b sono denominati, rispettivamente, "parte reale" e "parte immaginaria" del numero complesso ed i rappresenta l'unità immaginaria. Nel caso in cui a e b sono numeri interi si parla di "numeri complessi interi"; nel caso in cui a e b sono numeri in virgola mobile si parla di "numeri complessi in virgola mobile".

nelle colonne dalla 31 alla 50 e sono anch'essi giustificati a destra, se necessario; le colonne che non contengono dati devono sempre essere riempite con il carattere ASCII "SP";

- le stringhe di caratteri devono necessariamente essere racchiuse tra virgolette singole (ad esempio, nella *keyword* DATE la data deve essere specificata racchiudendola tra virgolette semplici, come in '10/12/2011'); se una stringa di caratteri contiene una virgoletta semplice, essa deve essere rappresentata mediante due virgolette semplici consecutive (ad esempio, 'nell'anno' deve essere espresso come 'nell''anno');
- tutte le stringhe devono contenere solo caratteri ASCII a 7 *bit*[44] e devono avere una lunghezza di almeno otto caratteri, per cui, se necessario, occorre inserire tanti caratteri ASCII "SP" di riempimento fino a raggiungere tale lunghezza; le stringhe possono contenere un massimo di 68 caratteri; i caratteri ASCII "SP" che precedono la stringa sono considerati parte del suo nome, mentre quelli che seguono la stringa sono considerati come caratteri di riempimento;
- tutti valori numerici, siano essi interi, a virgola fissa e a virgola mobile, con e senza segno, devono essere rappresentati con le equivalenti stringhe di caratteri ASCII: le variabili booleane sono rappresentate con i caratteri "T" e "F"; tutte le lettere utilizzate nelle forme esponenziali devono essere maiuscole (ad esempio, 1,2145E15).

Alcuni esempi di valori validi sono riportati **Tabella 4**[45].

Nome	Dimensione	Esempio
carattere ASCII	8 bit	'Saturn'
numero intero senza segno, un byte	8 bit	127
numero intero senza segno, due byte	16 bit	32767
numero intero senza segno, quattro byte	32 bit	1451342181
numero reale a singola precisione, notazione a virgola fissa	32 bit	3.14159
numero reale a singola precisione, notazione esponenziale	32 bit	0.314159E+01
numero reale a doppia precisione, notazione esponenziale	64 bit	0.3141592653525D+01

Tabella 4. Esempi di dati validi secondo la sintassi del formato FITS.

[44] Si veda la nota 30.

[45] L'esempio è tratto dal documento "*FITS File Format Summary*", estratto dalla *Encyclopedia of Graphics File Formats,* 2nd Edition, op.cit. Il documento è disponibile all'indirizzo http://www.fileformat.info/format/fits/egff.htm.

Struttura della Data Unit

La sezione *Data Unit*, che contiene i dati veri e propri, se presente segue immediatamente l'ultimo blocco da 2.880 *byte* della *Header Unit*. Come già visto, la presenza o l'assenza della *Data Unit* è indicata dai valori assunti dalle *keyword* NAXIS e NAXISn nella *Header Unit*.

Anche per la *Data Unit* valgono le regole di completamento a 2.880 *byte* viste per la *Header Unit*: se essa non raggiunge la dimensione di 2.880 *byte* o di un suo multiplo, vengono aggiunti, in coda ai dati, tanti caratteri ASCII "NUL" quanti ne sono necessari per raggiungere tale dimensione.

La *Primary Data Unit* contiene una matrice *n*-dimensionale (con *n* compreso tra 1 e 999) di dati che sono sempre riferibili alla rappresentazione digitale di una immagine[46] con dimensione compresa tra 1 e 999, la stessa tipologia di dati presenti in una *Image Extension HDU*. In questi due casi i valori contenuti nella *Data Unit* sono i valori dei *pixel* dell'immagine, e la tipologia di dato è specificata dalla *keyword* BITPIX nella *Header Unit* (si veda la **Tabella 3**): come già visto, sono consentiti sei tipi di dati: numeri interi senza segno ad 8 *bit*, numeri interi con segno a 16, 32 e 64 *bit*, numeri reali in virgola mobile a precisione singola a 32 *bit*

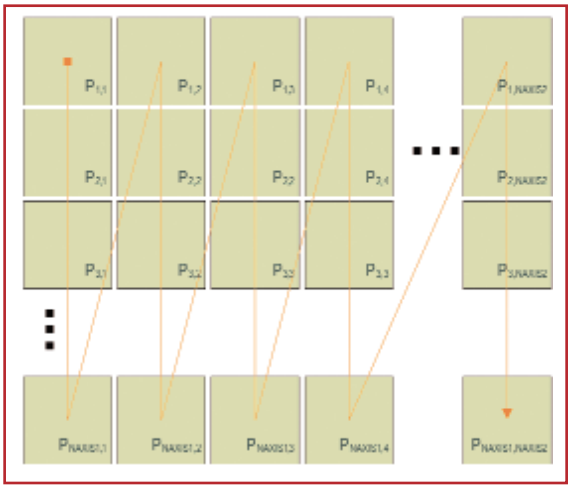

Figura 3. L'ordine seguito per definire i valori dei pixel nel caso di una immagine.

Figura 4. Esempio di allocazione dei valori dei pixel nel file FITS.

[46] Il formato FITS è stato utilizzato inizialmente per l'archiviazione di immagini, e questo rimane tuttora il suo utilizzo più diffuso. Tuttavia, il termine "immagine" va inteso non nel senso classico di immagine bidimensionale, ma in maniera più generale: infatti, il formato supporta matrici di dati di dimensione arbitraria (purché compresa tra 1 e 999); ad esempio può trattarsi di uno spettro 1-D (ad una dimensione), una immagine 2-D (bidimensionale), una immagine 3-D (tridimensionale), etc.

e a precisione doppia a 64 *bit* secondo la notazione ANSI/IEEE 754[47]. Tali valori vengono elencati l'uno di seguito all'altro seguendo la stessa convenzione adottata per le matrici multi-dimensionali nel linguaggio di programmazione Fortran[48]. Nel caso di una immagine bidimensionale, si inizia dal primo *pixel* in alto a destra ($P_{1,1}$), si prosegue procedendo lungo la "colonna" con quello immediatamente sotto ($P_{2,1}$), poi si passa al successivo ($P_{3,1}$) e così via fino ad arrivare al *pixel* $P_{NAXIS1,1}$. A questo punto si ritorna in alto al *pixel* $P_{1,2}$ nella colonna successiva, segue il *pixel* $P_{2,2}$, e così via fino ad arrivare all'ultimo *pixel* $P_{NAXIS1,NAXIS2}$ (si veda la **Figura 3**).

Ad esempio, nel caso di una immagine rappresentata dalla matrice a due dimensioni schematizzata in **Figura 4** (nella quale NAXIS = 2, NAXIS1 = 3 e NAXIS = 4), i valori dei singoli *pixel* vengono allocati nel *file* FITS iniziando dalla prima colonna (1, 5, 9), procedendo poi per la seconda (2, 6, 10), quindi per la terza (3, 7, 11) ed infine per la quarta (4, 8, 12).

La dimensione, N, espressa in *bit* di una immagine ad n dimensioni memorizzata in un *file* FITS può essere calcolata, escludendo i caratteri di riempimento, a partire dai valori delle *keyword* BITPIX, NAXIS e NAXISn presenti nella *Header Unit* mediante la seguente formula:

$$N = BITPIX \times (NAXIS1 \times NAXIS2 \times ... \times NAXISn)$$

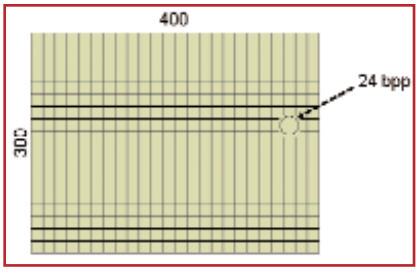

Figura 5. Esempio di immagine di dimensioni 300 x 400.

Ad esempio, nel caso di una immagine bidimensionale (NAXIS = 2) con profondità di colore pari a 24 (BITPIX = 24) e dimensioni di 300 x 400 *pixel* (NAXIS1 = 300 e NAXIS2 = 400), la dimensione in *bit* risulta (si veda la **Figura 5**):

$$N = 24 \times (300 \times 400) = 2.880.000 \text{ bit}$$

Esprimendo tale dimensione in *byte* (B) si ottiene:

[47] Lo standard IEEE 754 per il calcolo in virgola mobile (ANSI/IEEE 754-1985, IEEE *Standard for Binary Floating-Point Arithmetic* od anche IEC 60559:1989, *Binary floating-point arithmetic for microprocessor systems*) è lo standard più diffuso per il calcolo in virgola mobile. Definisce quattro possibili formati per la rappresentazione dei numeri in virgola mobile: a precisione singola (32 *bit*); a precisione doppia (64 *bit*); a precisione singola estesa (≥ 43 *bit*), raramente usato; a precisione doppia estesa (≥ 79 *bit*), supportata solitamente con 80 *bit*. La precisione singola è il minimo richiesto dallo standard, le altre sono opzionali.

[48] ISO/IEC 1539-1:2010 "*Information technology — Programming languages — Fortran — Part 1: Base language*", ISO/IEC, Ginevra, International Organization for Standardization.

N = 360.000 byte ~ 352 KiB.

Gli altri due tipi di estensioni standard, le *ASCII Table Extension HDU* e le *Binary Table Extension HDU*, contengono informazioni tabellari organizzate in righe e colonne. In generale, le tabelle binarie sono più compatte rispetto alle tabelle ASCII e risultano più veloci sia in fase di lettura che di scrittura perché i dati numerici sono memorizzati in una rappresentazione binaria leggibile direttamente dalla macchina. Tutte le voci all'interno di una data colonna di una *ASCII Table Extension HDU* o di una *Binary Table Extension HDU* corrispondono allo stesso tipo di dati. I formati di dati permessi per una colonna di una *ASCII Table Extension HDU* sono: numero intero, numero in virgola mobile a precisione singola o doppia, stringa di caratteri. Le *Binary Table Extension HDU* consentono anche l'inserimento di valori booleani ("T" o "F"), numeri binari e formati di dati complessi.
Per quanto riguarda l'utilizzo di algoritmi di compressione, lo standard FITS al momento non prevede l'impiego; pertanto i dati contenuti nella *Data Unit*[49] sono non compressi.

Alcuni esempi

Facciamo ora alcuni esempi per chiarire meglio quanto fin qui esposto. Consideriamo una immagine bidimensionale di dimensione 300 x 400 nella quale i *pixel* sono codificati con 16 *bit* (ovvero, con una profondità di colore pari a 16). In questo caso la *Header Unit* conterrà i seguenti *keyword record* (nota: le prime due linee servono solo per fornire informazioni sulla posizione ma non sono incluse nel *file* FITS):

```
   1    2    3    4    5    6    7    8
12345678901234567890123456789012345678901234567890123456789012
34567890123456789 0
```

```
SIMPLE =          T / il file È conforme allo standard FITS
BITPIX =         16 / numero di bit per ciascun pixel
NAXIS  =          2 / numero di assi
NAXIS1 =        300 / lunghezza del primo asse
NAXIS2 =        400 / lunghezza del secondo asse
END
```

[49] È stata proposta alla comunità FITS un'estensione allo standard FITS per consentire l'utilizzo della compressione; pertanto future versioni dello standard FITS potrebbero includere la compressione dei dati.

L'esempio appena visto fa uso delle sole *keyword* obbligatorie. A queste possono far seguito altre *keyword*, opzionali, che vengono utilizzate per meglio descrivere vari aspetti dei dati (come la data e l'orario dell'osservazione nel caso di dati astronomici). Frequentemente sono aggiunte le *keyword* ORIGIN, AUTHOR, COMMENT e HISTORY per documentare più dettagliatamente i dati.

Vediamo ora un esempio di *file* FITS che contiene la rappresentazione digitale di una immagine di dimensioni 1000 x 1500 con una profondità di colore pari a 32. In questo caso il *file* è costituito solamente da una *Primary Header Unit* (che fa uso anche di *keyword* opzionali) ed una *Primary Data Unit*.

Si noti che il blocco di caratteri (apparentemente incomprensibile) che inizia immediatamente dopo il *keyword record* END altro non è che la rappresentazione in caratteri ASCII della *Data Unit* (che, ovviamente, è costituita da una sequenza di *bit*). I *keyword record* possono comparire in qualsiasi ordine al-

```
      1       2       3       4       5       6       7       8
12345678901234567890123456789012345678901234567890123456789012345678
901234567890

SIMPLE  =                    T / il file e' conforme allo standard FITS
BITPIX  =                   32 / numero di bit per ciascun pixel
NAXIS   =                    2 / la tabella è una matrice 2D
NAXIS1  =                 1000 / lunghezza del primo asse
NAXIS2  =                 1500 / lunghezza del secondo asse
DATE    = '02/12/2011'         / data di creazione
ORIGIN  = 'Digitalia'          / FITS file originator
AUTHOR  = 'Mario Rossi'        / creatore
OBJECT  = 'Manoscritto_1'      / nome dell'oggetto osservato
HISTORY  'prima versione'
COMMENT  'Esempio di Header Unit'
END
?„»?Šó□?□Ù□>Đó¢?h?qß<?□•^?~É□?¯U×>+-$¾Û<j=š □?□Çô?>z>Ò"Ô>Ù¼ë>*-
O?%8¤?)£>□šM?□+?ˆ?\?7"=?#;Ç?□□i>¦?†‹ê?g!ß?j^À?q□?B□~?\[□?'´¤?¤½Þ? @[
?±I=?o|›?<ˆ?è_4?Ý®à@□?P□K?<ƒ<?\>x?KDÏ?zØi?Ã##?¸‹□?¸3\?ºäP@h□t@E£□
@RË^@y>O@§¦,@i□□@ĐµÕ@òhmA□éq@ÚÉ•@Ú¼sA7ôõAzb□AfxFAV9LAD[¿Aó1A,
˜+A2□I›A4ò'A}□(A±¬›A§TÚAq³AHjZAMe+A}¬AŠéA□□$A¯□óA³/□AÀA□AÑúÂAë•|B
(LÍB1□˜B)ä£B□□AûTÇA¥,úAKE¦A.0¢A□ß,A□ö`@â¤Ï@äZ@¨Áš?¢□?L™š>_u'?;dâ>®
á+>□RT>ãr"?Ké?D"□?□º¢>ß□3?V™8?Q!?□¦+?ë[>mëw?□<ì?¬>X?□v?bœÔ>à3?˜?;
□‰?is?□□»?ºé³?□'+?›m?gt1>ËÉS><pw?GXt?□#4? #;?ºâ7?¿8?1X0?DRç?Œ¸b?ÜkV?Ë
uÔ?¬(™?Šùþ?œŸ¬?ê□Ò?äy:?Ì~¦?ç¿□?Éì|?¹dÐ?ÀÍƒ?¬|p@Ÿ□@5ü?èõsdjklaakkal
[...]
Ë±@□Ôà@Q£@'Œ@+vì@šÈÝ@iÒ□@¨□Œ@œ§9@±åÝ@è"=@ë□Ê@þ´@áq|A□ÀA'¬
œA+$ÀAËFÂA¿+õA˜š¨A□ÈA^□ßA²¿oAÍ□±Añ:IB $B□…□B□™7B□„õB+>iB\^□B¦B{Ú
3B□þ□BÙ□A¯□,A^dKAd°AÑoAã9@óu¸A□¼Ÿ@Ý;!?ß?Þº□?Hj…?q³Ë>Ð□©?šDÃ?Xaá
@□□ô>àü´?¨^Å?À□ü>h³?Š□^>Û^Y>£¶Ã?□Ò□>Ð5»?_„□?ˆoÁ?‰‰?Y,□?'Å>õˆ
¨?)□Ö?ij*?«j2?¢¸g?ƒàé?¥□□?□©Ö?Aˆ-?AÃ□?Ü×ò?ºÐG?©®ü?□›V?□□□?›ûÃ@□¹□
?Ïùê?7Š□?Œál?Yjó?³zW>ëL:=ƒK□?kJç?£□S?Ø ?¢?\?†°1?"RÍ?»•Á?Ï‰"?µ¾¤?·□4?"ih
?¸RD?+Û¨?…v¦?$ëá?ÒÏd?Óqf?Õã$?õòÕ?¿± ?£□m?ãR¿½;Đ1?‹¸‹?¯¾p?¸ƒ•?+tO@□-Ý
>Î\¯?{&Î?□□??90Ü?□‹?‰o0?□$ô>-□}?«6¶?µçÆ?˜+"?³×^@1¬m>,#½?,Ìi>Y¦G?ºü
ä?□ÃÒ?IÆt?„ò?P®□?<□²?□é#?□□W?P`{>ÚÒ7>«Þi?,Ýi?□&□?n+µ>‰c3><|¯>èW
`>²êw?,³Ô?Y³>□xB?[íG?□qy>ù|>?H…t>å8|¾Ê÷¨¿F[>:¿~?•yt?Ëxò?Kö□?!□O?tâN?8
D¼¿Ç½ÿ+8?Œ
```

l'interno della *Header Unit* e sono previste poche eccezioni: la *keyword* SIMPLE deve essere sempre la prima, seguita dalla *keyword* BITPIX come seconda e dalla *keyword* NAXIS come terza; la *keyword* END deve essere sempre l'ultima. Vi sono molte altre regole[50] che disciplinano il formato esatto di ciascuna *keyword* per cui è generalmente meglio affidarsi a software di interfaccia per costruire correttamente o analizzare le *keyword* piuttosto che scrivere o leggere direttamente un *file* FITS. Nonostante ciò è sicuramente possibile "lavorare" anche a mano, utilizzando un semplice *editor* di testo, e questa possibilità costituisce uno dei punti di forza del formato perché lo rende molto trasparente.

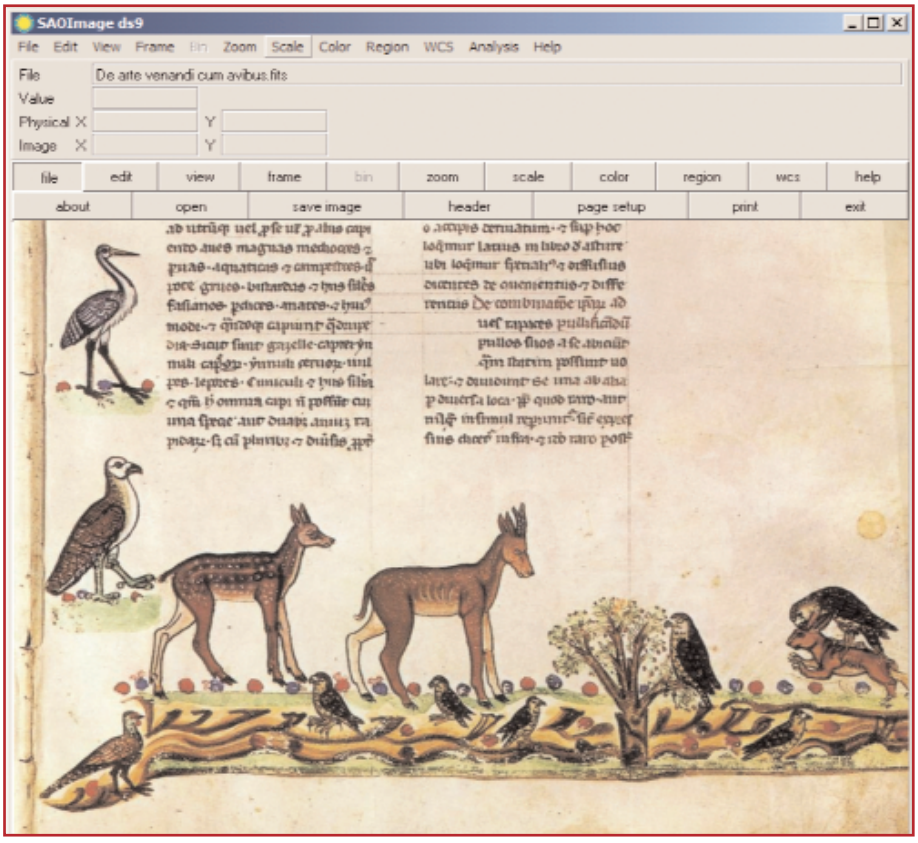

Figura 6. Una pagina del De arte venandi cum avibus in formato FITS visualizzata con SAOImageDS9 (fonte: immagine di pubblico dominio disponibile all'indirizzo http://it.wikipedia.org/wiki/De_arte_venandi_cum_avibus)

[50] Si veda la specifica della versione 3.0 del formato (*Definition of the Flexible Image Transport System (FITS). The FITS Standard. Version 3.0*, cit.).

Produzione e fruizione delle immagini in formato FITS

Per creare e leggere *file* in formato FITS è possibile utilizzare una folta serie di applicativi software[51]. Alcuni di questi sono *plug-in* che vanno ad estendere le funzionalità di altri programmi, come nel caso di *FITS Liberator*, un *plug-in* per Adobe Photoshop, grazie al quale quest'ultimo acquisisce la capacità di visualizzare immagini in formato FITS. Altri programmi sono, invece, di tipo *stand-alone*, ovvero non richiedono l'installazione di altre applicazioni o di *file* di supporto.

Tra i software che possono essere utilizzati liberamente per la fruizione di *file* codificati nel formato FITS va segnalato Gimp (*GNU Image Manipulation Program*)[52], probabilmente il più noto programma *open source* di fotoritocco, che permette di leggere e scrivere nativamente *file* nel formato FITS. Altri visualizzatori gratuiti (ma non *open source*) sono *XnView*[53] e *IrfanView*[54]. Questi programmi sono in grado di leggere semplici *file* FITS consententi immagini, ma spesso non sono in grado di interpretare quelli contenenti dati più complessi, come le tabelle binarie. Tra gli applicativi per la gestione di *file* FITS più completi va segnalato *SAOImage DS9* (si veda la **Figura 6**) realizzato dallo *Smithsonian Astrophysical Observatory* (SAO); è di tipo *stand-alone* (non richiede installazione o *file* di supporto) ed è disponibile per quasi tutti i principali sistemi operativi (Solaris, Linux, MacOSX e Windows).

Vi sono numerose altre applicazioni in grado di gestire, con funzionalità più o meno complesse, i *file* in formato FITS. Sul sito del *High Energy Astrophysics Science Archive Research Center* (HEASARC), il principale archivio di immagini acquisite durante le missioni stazionali della NASA, è disponibile una nutrita serie di strumenti software[55]. Il principale è *FTOOLS*[56], un pacchetto molto completo di programmi per manipolare e analizzare i dati in formato FITS. I programmi più comunemente utilizzati sono disponibili anche singolarmente, senza dover installare l'intero pacchetto[57].

[51] Un esaustivo elenco di applicativi e librerie che sono in grado di leggere e scrivere *file* nel formato FITS è disponibile sul sito del *FITS Support Office at NASA/GSFC*, all'indirizzo http://fits.gsfc.nasa.gov/fits_libraries.html. Altre informazioni possono essere desunte dall'elenco compilato dall'*Astronomy Education Committee*, disponibile all'indirizzo http://www.phy.duke.edu/~kolena /imagepro.html.

[52] Il programma è disponibile liberamente all'indirizzo principale del progetto GIMP http://www.gimp.org, oltre che da numerosi altri siti.

[53] *XnView* è un visualizzatore di immagini gratuito con funzioni di fotoritocco che consente anche di convertire i *file* in diversi formati (il programma supporta più di 400 formati differenti ed effettua la conversione verso circa 50 formati diversi). È disponibile all'indirizzo http://www.xnview.com.

[54] Si veda il sito *web* http://www.irfanview.com.

[55] Si veda all'indirizzo http://heasarc.gsfc.nasa.gov/docs/software.html.

[56] Si veda all'indirizzo http://heasarc.gsfc.nasa.gov/docs/software/ftools/ftools_menu.html.

[57] Tra di essi troviamo *fcopy*, un potente programma per copiare selettivamente parti di un *file* FITS in un nuovo *file*; *flist*, per stampare il contenuto (*Header Unit*, immagini, tabelle, etc.) di un *file* FITS in un *file* di testo ASCII; *fverify*, per verificare se un *file* è conforme alle specifiche dello

Accanto a FTOOLS va segnalato *FITS Tools*[58], un pacchetto più compatto di programmi di utilità per visualizzare o modificare il contenuto dei *file* FITS. È scritto nel linguaggio di programmazione C utilizzando la libreria CFITSIO[59] e il codice sorgente può essere modificato e riutilizzato da chiunque. Occorre notare che frequentemente i gruppi di ricerca scientifici scrivono il proprio codice per interagire con i propri dati FITS, utilizzando gli strumenti disponibili nel linguaggio di programmazione scelto. Questo è possibile grazie al fatto che la maggior parte dei linguaggi di programmazione utilizzati nei lavori scientifici (tra cui il C, il FORTRAN, Java, Perl, PDL, Python, R, S-Lang e IDL) prevedono un ampio supporto al formato FITS. Sono disponibili anche numerose applicazioni software in grado sia di visualizzare le immagini contenute nei *file* FITS che di convertirle in altri formati per immagini (ad esempio, TIFF; JPG, BMP, TGA, etc.)[60]. Va ricordato che FITS è un formato di dati molto generale e viene utilizzato per tipologie di dati (sia nel campo astronomico che in altri campi) a volte molto differenti, per cui non tutti gli applicativi sono necessariamente in grado di leggere ogni tipo di *file* FITS.

Il formato FITS come formato per la conservazione

Come si sarà potuto osservare dalla lettura di quanto fin qui riportato, il formato FITS gode di molte proprietà che rendono interessante la sua adozione come formato per la conservazione nel tempo delle immagini ottenute mediante processi di digitalizzazione. Innanzitutto, va fatta rilevare la sua estrema semplicità: sostanzialmente un *file* in formato FITS è costituito da due componenti principali, la *Header Unit* contenente i metadati e la *Data Unit* contenente i dati veri e propri. Questa organizzazione può essere ripetuta in linea di principio senza limiti (non c'è limite al numero di *Extension HDU* che possono essere aggiunte) rendendo il formato scalabile. Sia le *Header Unit* che le *Data Unit* sono, a loro volta, composte da blocchi logici che devono avere la dimensione di 2.880 *byte* o suoi multipli, per cui se l'ultimo blocco logico di una unità non raggiunge tale dimensione devono essere aggiunti tanti caratteri di riempimento (il carattere ASCII "SP" nel caso delle *Header Unit* o il carattere ASCII "NUL" nel caso delle *Data Unit*) fino ad arrivare alla dimensione esatta di 2.880 *byte* o di un suo multiplo.
A sua volta ogni blocco logico contiene 36 record, chiamati *keyword record*, ciascuno di dimensione 80 byte, eventualmente completato con caratteri di riempi-

standard FITS; *fv*, un visualizzatore ed editor di *file* FITS con interfaccia grafica (quest'ultimo è disponibile all'indirizzo http://heasarc.gsfc.nasa.gov/docs/software/ftools/fv/ fv.html).

[58] Si veda all'indirizzo http://heasarc.gsfc.nasa.gov/docs/software/fitsio/cexamples.html.

[59] Si veda all'indirizzo http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html.

[60] Un elenco completo di applicativi software in grado sia di visualizzare le immagini che di convertirle in altri formati per immagini è riportato all'indirizzo http://fits.gsfc.nasa.gov/fits_viewer.html.

mento per raggiungere tale lunghezza. Come si può osservare, si tratta di un organizzazione logica estremamente lineare.
Vediamo ora più in dettaglio quali sono i requisiti che il formato FITS possiede: condurremo l'analisi sulla base della classificazione dei requisiti (si veda la **Figura 7**) elaborata nel corso di uno studio condotto presso il Laboratorio di Informatica Documentale dell'Università degli Studi di Macerata, che ha permesso di enucleare i requisiti che un formato elettronico deve possedere per poter essere considerato compatibile con un processo di conservazione digitale[61].

Figura 7. I requisiti desiderabili per un formato compatibile con un processo di conservazione digitale (fonte: S. Pigliapoco, S. Allegrezza, Produzione e conservazione del documento digitale. Requisiti e standard per i formati elettronici. Volume I, Edizioni EUM, Macerata, 2008).

Analizziamo dapprima il formato con riferimento ai requisiti generali di *primo livello* (ovvero, quei requisiti che devono essere presi in considerazione in prima battuta per operare una selezione iniziale):

– *Non proprietà*. Il formato FITS è gestito dalla comunità scientifica internazionale (in particolare, quella degli astronomi) ed è, come tale, non proprietario; il mantenimento delle specifiche è attualmente a carico dello *IAU FITS Working Group*.
– *Apertura*. Il FITS è un formato aperto, essendo basato su una specifica, il *FITS Reference*, disponibile pubblicamente. Non vi sono *royalty* (né mai ci saranno) da pagare per l'utilizzo del formato, per cui chiunque può creare applicazioni in grado di gestire i *file* FITS.

[61] I risultati di tale studio sono stati pubblicati nel volume S. Pigliapoco, S. Allegrezza, *Produzione e conservazione del documento digitale. Requisiti e standard per i formati elettronici. Volume I*, Edizioni EUM, Macerata, 2008.

- *Completa documentazione*. Il formato FITS è pienamente documentato nel *FITS Reference*, liberamente disponibile.
- *Standardizzazione*. Il FITS è un formato riconosciuto *standard* fin dal 1993 (si veda il paragrafo 4). Si noti che il FITS è uno *standard aperto* (*open standard*), essendo un formato che possiede entrambe le proprietà di standardizzazione ed apertura.
- *Ampia adozione.* Il formato FITS è attualmente il formato più utilizzato nel campo dell'astronomia. Esistono, inoltre, numerose applicazioni per la creazione e la visualizzazione di *file* in formato FITS nonché per la loro conversione da e verso altri formati per immagini.
- *Trasparenza.* Una delle caratteristiche principali del formato FITS è costituita dal fatto che i metadati vengono memorizzati in formato ASCII *human readable*, in modo che chiunque possa facilmente esaminare, mediante un semplice *editor* di testo, la *Header Unit* del *file* per poter avere informazioni sul *file* stesso. Inoltre, le informazioni relative ai *pixel* delle immagini (contenute nella *Data Unit*) sono memorizzare in modo molto semplice, senza alcuna compressione; grazie alla sua semplicità è possibile, e lo sarà nel futuro, leggere un *file* in formato FITS utilizzando strumenti software di base.

Analizziamo ora il formato con riferimento ai requisiti generali di *secondo livello* (ovvero, quei requisiti, non meno importanti dei primi, ma che vanno presi in esame successivamente dopo aver selezionato i formati sulla base dei requisiti di primo livello):

- *Robustezza*. Grazie alla sua semplicità e al fatto che non è compresso, la corruzione di alcune sequenze di *bit* non produce di norma la perdita del contenuto informativo del *file* e il degrado dell'immagine si mantiene generalmente entro livelli accettabili.
- *Stabilità*. Il formato FITS è molto stabile: nell'arco di oltre trenta anni si sono succedute solo tre versioni.
- *Compatibilità all'indietro e in avanti*. Il formato è compatibile all'indietro e in avanti. FITS è stato progettato tenendo ben in mente l'aspetto della conservazione a lungo termine: non a caso la massima "*once FITS, always FITS*" che è stata coniata nei suoi confronti stabilisce con chiarezza che non è permessa alcuna modifica al formato che renda non più valido un *file* FITS già esistente.
- *Auto-contenimento*. Tutte le informazioni necessarie per poter rappresentare correttamente (a video o a stampa) un *file* FITS, sono incluse al suo interno e non sono previsti, ad esempio, *link* (collegamenti) verso risorse esterne.
- *Auto-documentazione.* Un *file* in formato FITS è auto-documentato. Le *keyword* che compaiono nella *Header Unit* hanno lo scopo di documentare esaustivamente il *file* e forniscono informazioni quali la dimensione, le origini, la storia dei dati e quant'altro il soggetto produttore desideri inserire.

Inoltre, oltre alle *keyword* standard è possibile utilizzare anche altre *keyword* coniate appositamente per meglio descrivere la particolare tipologia di dato che si intende memorizzare nel *file*[62].

- *Indipendenza dal dispositivo*. Essendo un formato multi-piattaforma, un *file* in formato FITS può essere visualizzato, stampato o comunque riprodotto in maniera affidabile e coerente, sempre nello stesso modo indipendentemente dalla piattaforma hardware e software utilizzata.
- *Assenza di meccanismi tecnici di protezione.* Il formato FITS non prevede meccanismi di protezione tecnica. Non è prevista, ad esempio, la crittografia così come il controllo dell'accesso mediante *username* e/o *password*[63].
- *Assenza di limitazioni sull'utilizzo*. Il formato FITS non presenta limitazioni di alcun genere sul suo utilizzo né attuale né futuro.
- *Accessibilità*. È possibile definire delle *keyword* per definire del "testo alternativo" che può essere letto dal software di visualizzazione per descrivere l'immagine; ciò si rivela molto utile nel caso in cui l'immagine sia fruita da soggetti con disabilità visive
- *Non modificabilità*. Il FITS può essere modificato ricorrendo ad apposite applicazioni software (FITS *editor*). Tuttavia, qualora necessario, la sua non modificabilità può essere assicurata dall'utilizzo di accorgimenti tecnici quale l'utilizzo di *checksum* o di firme elettroniche, etc.
- *Sicurezza*. Allo stato attuale delle conoscenze, il FITS non può contenere virus o altre forme di *malware*.
- *Efficienza*. Il formato FITS è di tipo non compresso, pertanto le dimensioni dei *file* codificati in questo formato sono spesso piuttosto elevate.

Venendo ora ai requisiti *specifici* della particolare categoria a cui il formato FITS appartiene (ovvero, i formati per immagini) e che prendono in considerazione le proprietà che si ritengono importanti per quei formati ai fini della migliore conservazione nel tempo, osserviamo che esso gode di altre interessanti proprietà non sempre riscontrabili nei formati per immagini che normalmente si utilizzano per scopi analoghi.

Innanzitutto il formato FITS consente di memorizzare immagini assolutamente fedeli rispetto all'originale, sia per quanto riguarda la risoluzione che la profondità di colore. Infatti, teoricamente non vi sono limiti alla risoluzione delle immagini che

[62] Nel caso del progetto di digitalizzazione dei manoscritti della Biblioteca Apostolica Vaticana, nella *Header Unit* sono stati specificati oltre 900 *keyword record* contenenti informazioni relative ai parametri di acquisizione e di catalogazione (segnatura, formato, etc.) secondo lo standard METS (*Metadata Encoding and Transmission Standard*).

[63] Nel caso sia necessario un controllo dell'accesso, esso può essere realizzato mediante meccanismi esterni: ad esempio, è possibile comprimere con l'algoritmo ZIP il *file* FITS ed applicare al *file* così ottenuto la protezione mediante *username* e *password*.

si possono memorizzare[64]: la Biblioteca Apostolica Vaticana sta utilizzando apparati di acquisizione ad elevatissima risoluzione (50 Mpixel), ma se in futuro dovessero essere disponibili scanner e macchine fotografiche con risoluzioni maggiori (ad esempio, 500 Mpixel) non ci sarebbero difficoltà a continuare ad utilizzare il formato FITS.

Inoltre, FITS consente di conservare intatta anche la qualità dei colori, elemento di grande valore per gli studiosi, poiché è possibile specificare profondità di colore anche di 64 *bit* (equivalente ad un numero di colori visualizzabili pari a oltre 18 miliardi di miliardi)[65]: di fatto si tratta di una profondità di colore quasi illimitata e certamente di gran lunga superiore alla capacità dell'occhio umano di distinguere colori con differenze così impercettibili. In questo modo è possibile memorizzare immagini senza alcuna perdita di colore (si pensi ai colori delle miniature e al colore oro, estremamente difficile da riprodurre) e con la certezza che i colori memorizzati saranno fedeli agli originali.

Infine, contrariamente ad altri formati immagine, con il formato FITS è possibile memorizzare più immagini nello stesso *file*, per cui sarebbe possibile, in linea di principio, memorizzare un intero manoscritto in u unico *file*.

Gli altri formati proposti per la conservazione nel tempo delle immagini non sempre soddisfano un numero così elevato di requisiti. Ad esempio, il formato immagine TIFF (*Tagged Image File Format*), utilizzato per molti anni sia da enti pubblici che da aziende private per conservare negli archivi digitali ogni genere di documentazione (fatture, corrispondenza, contratti, etc.), non è dotato degli stessi requisiti del FITS (pur rimanendo tuttavia un valido formato).

Conclusioni

Come si è potuto vedere, FITS è un formato che soddisfa un gran numero di requisiti desiderabili per un formato compatibile con un processo di conservazione digitale; inoltre, è uno dei pochi formati che ha superato la prova del tempo essendo, a distanza di oltre trenta anni dalla sua nascita, tuttora ampiamente adottato. Quindi si può ragionevolmente affermare che possiede tutti i requisiti per proporsi come formato consigliabile per la conservazione a lungo termine delle im-

[64] L'unico limite è rappresentato da eventuali limitazioni imposte dalla piattaforma hardware e software che dovrà poi gestire tali file; ad esempio, è noto che alcuni *file system* impiegati nei sistemi operativi impongono limitazioni sulle dimensioni massime dei *file*, come nel caso dei 4GB massimi imposti dal *file system* FAT32.

[65] Si ricorda che la profondità di colore va specificata come *value* della *keyword* BITPIX e può assumere valori i valori riportati in **Tabella 3**. Nel caso si scelga un valore pari a 16, la profondità di colore è pari a 2^{16} = 65.536; nel caso si scelga, invece, un valore pari a 32, la profondità di colore è pari a 2^{32} = 4.294.967.296; infine, scegliendo un valore pari a 64, la profondità di colore è pari a 2^{64} = 18.446.744.073.709.551.616, vale a dire oltre 18 miliardi di miliardi di possibili colori. Il progetto di digitalizzazione dei manoscritti della Biblioteca Apostolica Vaticana prevede l'utilizzo di una profondità di colore pari a 32 *bit*.

magini acquisite mediante progetti di digitalizzazione. È per questo che occorre cominciare seriamente ad interrogarsi, sia prima di avviare nuovi progetti di digitalizzazione che in previsione delle operazioni di riversamento sostitutivo che nel tempo dovranno essere effettuate sul materiale già digitalizzato, se non sia conveniente incominciare ad utilizzare questo formato in luogo degli altri formati tipicamente adottati (come il TIFF, il BMP, il JPG, etc.). Non è escluso che il formato FITS diventi lo standard del futuro per la conservazione nel tempo dei documenti provenienti da digitalizzazioni ed anzi sembra che ci sia l'intenzione di proporre all'*International Organization for Standardization* (ISO) il suo riconoscimento come formato standard per la conservazione digitale di libri e manoscritti antichi. Se ciò si avvererà, si avrà, finalmente, a disposizione un formato "ufficiale" che tutte le biblioteche e gli archivi del mondo potranno utilizzare per preservare i patrimoni digitali ottenuti da processi di digitalizzazione.

The Vatican Apostolic Library has begun a massive digitization project of over eighty thousand precious manuscripts; has decided to use the FITS format for the long-term conservation of acquired images, instead of the usual image formats (TIFF, BMP, JPG, etc.).
At first sight this choice might seem unusual, since the FITS format has been used in astronomy for over thirty years, but almost certainly never for long-term archiving and conservation of images acquired by digitization.
However, while analyzing the definition of the format it was discovered that it possesses hallmarks that render it fully compatible with the process of digital conservation.
Starting from the digitization project of the Vatican Apostolic Library, this paper aims at analyzing the characteristics of the FITS format in an attempt to determine whether it can aspire to become the "standard format" to be adopted in digitization projects, so that libraries and archives can ensure a long-term conservation of digital heritage, already acquired or to be acquired.