

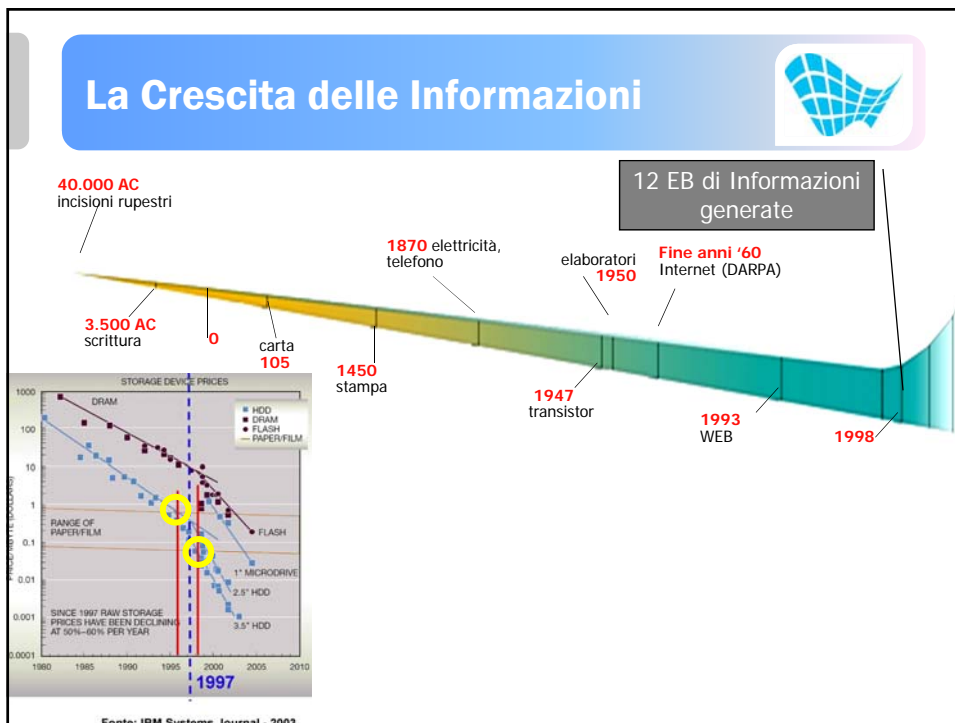
**"ICT & PMI: senza carta con le nuove norme"**  
a cura della  
**Associazione ICT Dottori Commercialisti**

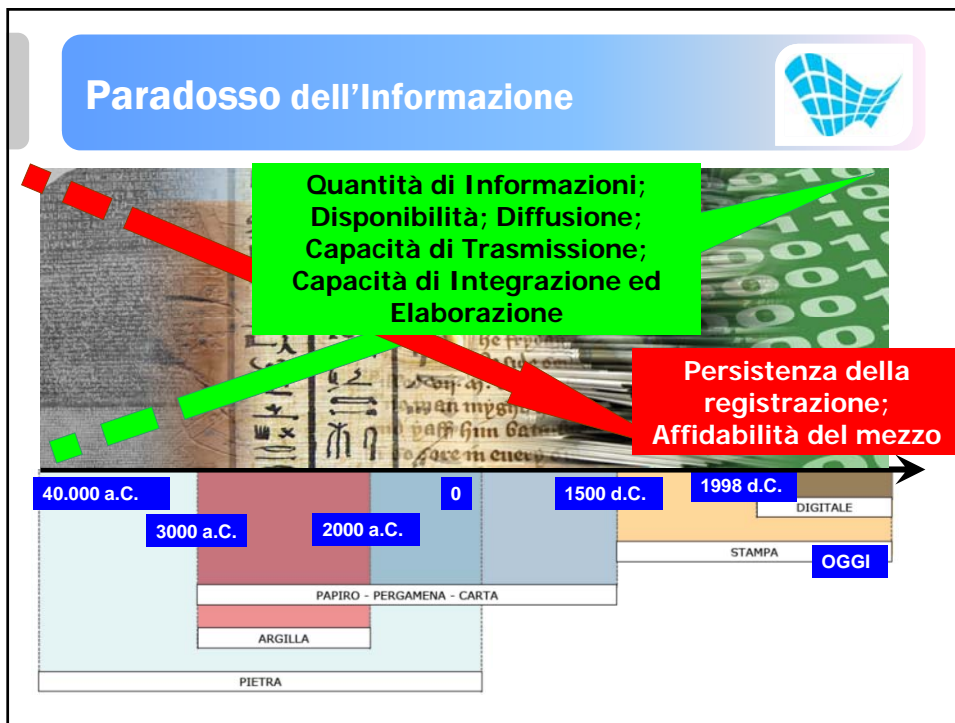
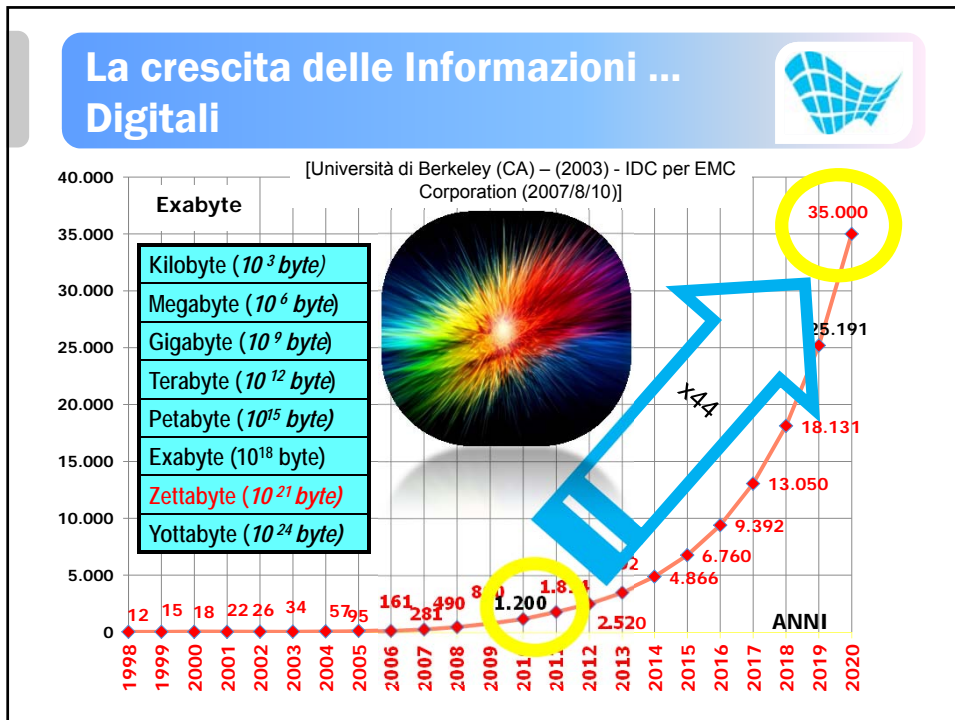


**I Formati di Conservazione  
dei Documenti d'Impresa**



**Vincenzo Gambetta**  
(Torino, 25 novembre 2011)





**Alcune considerazioni sulla tipologia di documento**

Documento Tradizionale	Documento Informatico/Digitale
"Immodificabile"	Modificabile/Elaborabile (Esigenza di tecnologie che ne assicurino una qualche forma di immodificabilità)
"Esigenza di un Supporto fisico". Inscindibilità tra Contenuto e informazioni correlate dal Supporto	Supporto fisico <b>IRRILEVANTE</b> se non ai fini della sua conservazione o trasporto Contenuto e informazioni correlate <b>indipendenti dal Supporto</b>
Visualizzazione implicita nella fisicità	Necessità di una tecnologia che lo visualizzi
Sottoscrizione e datazione legate al "contenuto" dal Supporto	Modalità di datazione e sottoscrizione legate a specifiche tecnologie
Modalità di attribuzione e datazione possono essere dedotte dalla fisicità del supporto (grafia, analisi dell'inchiostro e delle caratteristiche del supporto, ...)	

**Alcune considerazioni sulla tipologia di documento**

Documento Tradizionale	Documento Informatico/Digitale
"Immodificabile"	Modificabile/Elaborabile (Esigenza di tecnologie che ne assicurino una qualche forma di immodificabilità)
"Esigenza di un Supporto fisico". Inscindibilità tra Contenuto e informazioni correlate dal Supporto	Supporto fisico <b>IRRILEVANTE</b> se non ai fini della sua conservazione o trasporto Contenuto e informazioni correlate <b>indipendenti dal Supporto</b>
Visualizzazione implicita nella fisicità	Necessità di una tecnologia che lo visualizzi
Sottoscrizione e datazione legate al "contenuto" dal Supporto	Modalità di datazione e sottoscrizione legate a specifiche tecnologie
Modalità di attribuzione e datazione possono essere dedotte dalla fisicità del supporto (grafia, analisi dell'inchiostro e delle caratteristiche del supporto, ...)	

**THINK DIGITAL**

## Formazione del Documento Informatico

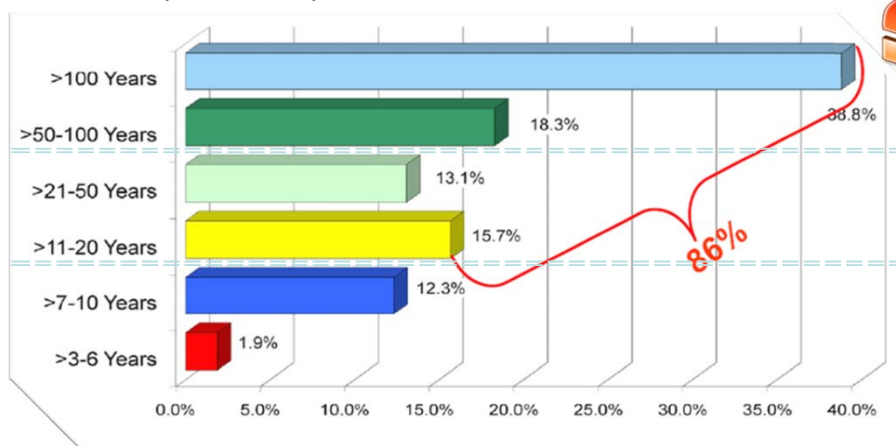


1. Redazione tramite l'impiego di software
2. Acquisizione
  - della copia per immagine su supporto informatico di un documento analogico,
  - della copia informatica di un documento analogico,
  - per via telematica o su supporto informatico;
3. Registrazione informatica delle informazioni risultanti da
  - transazioni informatiche o
  - dalla presentazione telematica di dati attraverso moduli o formulari;
4. Generazione o raggruppamento, anche in via automatica, di un insieme di dati, provenienti da una o più basi dati secondo una struttura logica predeterminata e memorizzata in forma statica.

## Conservare a lungo termine?



- Per quanto tempo dovete conservare i vostri dati?



Fonte: SNIA - 100 Year Archive Requirements Survey - Gennaio 2007

## Tempi di conservazione obbligatori per la documentazione di business



	Anni di Conservazione			
	IVA	Imposte Dirette	Esigenze contabili	Esigenze Comm. II
Austria	Da 7 a 22	7	7	7
Belgio	7	5	10	10
Danimarca	Da 5 a 10	5	5	5
Finlandia	6	6	6	N/D
Francia	6	6	6	6
Germania	10	10	10	10
Irlanda	5	5	8	N/D
<b>Italia</b>	<b>4</b>	<b>4</b>	<b>10</b>	<b>10</b>
Olanda	Da 7 a 9	7	N/D	7
Norvegia	10	10	10	10
Portogallo	10	10	10	10
Spagna	4	4	6	6
Svezia	7	7	10	N/D
Svizzera	Da 10 a 20	10	10	10
Regno Unito	6	6	6	N/D

## La Conservazione della Memoria Digitale

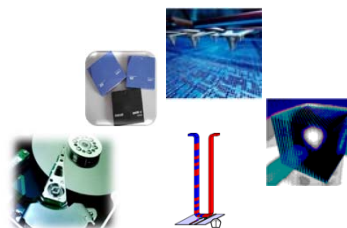


### Formato



**Metadati e Regole per gestire l'intero Ciclo di Vita delle Informazioni**

### Consultabilità, Autenticità, Integrità e Affidabilità



**Integrità delle registrazioni fisiche**

## I Formati



- **“Formato”** - modo in cui l’informazione digitale è strutturata e registrata sul supporto di memorizzazione
  - **Corpo**
  - **Testata o header**
  
  - Formati preferenziali
  - Formati compatibili
  - Portabilità
  - Formati proprietari e formati standard

## I Formati



- Formati → in aumento (programmi/piattaforme)
- ~ **1.500** formati (estensioni); per un terzo non è disponibile la descrizione
  - **Testi/documenti** - (DOC, DOCX, XML, HTML, PDF,...)
  - **Immagini** - (GIF, JPG, BMP, TIF, CGM, EPS, SVG, ...)
  - **Suoni** - (MP3, WAV, ...)
  - **Filmati** - (MPG, MPEG, AVI, WMV,...)
  - **Eseguibili** - (EXE, ...)
  - **Calcolo** - (XLS, MDB, ...)
  - **Archiviazione e Compressione** - (ZIP, RAR, ...)
  
  - **Dati binari**, più o meno strutturati, come, ad esempio, gli output di strumentazione di laboratorio
  
  - **E-mail** (SMTP/MIME, ...)

## I Formati e l'operatività



- Vanno valutati in funzione de
  - La **diffusione** - il numero di persone ed organizzazioni che li adotta
  - La **portabilità** - usabilità su piattaforme differenti (meglio se essa è indotta dall'aderenza a standard documentati e accessibili)
  - Le **funzionalità** che l'utente ha a disposizione per elaborare l'informazione e collegarla ad altre (ad esempio gestione di link)
  - La **capacità di occupare il minor spazio possibile in fase di memorizzazione**
  - La **capacità di gestire il maggior numero possibile di metadati durante tutto il ciclo di vita del documento**
  - La **capacità dei programmi di gestire contemporaneamente un numero congruo di formati** (in funzione delle esigenze dell'utente)
  - La **diffusione** dello stesso e dei visualizzatori che consentono la fruibilità delle informazioni
  - L'**apertura** - la conformità a specifiche pubbliche
  - La **sicurezza** - il grado di immodificabilità del contenuto del file e l'immunità dall'inserimento di «Codice Maligno»

## I Formati e la Conservazione nel Lungo Periodo



- Caratteristiche che un formato deve avere [per essere "riconosciuto" e utilizzato correttamente nel tempo e garantire interoperabilità]
  - **Immodificabilità** - il formato deve essere strumentale a far assumere al documento le previste caratteristiche di immodificabilità e staticità (a qualunque livello, compresa struttura e modalità di visualizzazione)
  - **Mantenimento della fedeltà di presentazione del contenuto** - ciò non significa semplicemente conservare nel tempo la struttura del documento, ma anche i cosiddetti "font" e tutte le altre caratteristiche utili alla completa comprensione del documento
  - **Non contenere riferimenti ad altre informazioni o documenti** (ad esempio link ipertestuali) non sotto il controllo della stessa "unità di informazione" o documento.
  - Il formato di conservazione **deve corrispondere fedelmente ad uno standard** (formale o de facto) **interamente descritto e le cui specifiche siano state rese di dominio pubblico senza limitazioni**; in subordine se è un formato proprietario devono essere pubbliche le specifiche tecniche
- Il processo di conversione dal formato di un documento "operativo" ad uno di conservazione deve essere semplice.

## I Formati e la Conservazione nel Lungo Periodo



- Alla data soddisfano i requisiti appena illustrati:
  - PDF/A, descritto dagli standard ISO 19005-1: 2005(PDF 1.4) e ISO 19005-2:2011 (PDF 1.7)
    - può disporre di uno "strato" XML (in grado di trasportare dati strutturati e metadati)
    - ha la possibilità di "inglobare" nativamente la Firma Digitale
    - Disponibilità di prodotti per verificare la conformità
  - Formato immagine (Raster) TIFF in versione non compressa o compressa senza perdita di informazione.
    - di questo formato vi sono parecchie versioni, alcune delle quali proprietarie
    - in genere le specifiche sono pubbliche e non soggette ad alcuna forma di limitazione.
    - sono formati molto utilizzati per la conversione in digitale di documenti cartacei.
    - Il loro impiego, se attuano una forma di compressione, va valutato attentamente.
  - JPEG 2000 (ISO/IEC 15444-1)
    - può utilizzare la compressione senza perdita di informazione
    - consente di associare metadati ad un'immagine.
    - diffusione limitata

## Altri Formati



- Open XML (OOXML)
  - Pubblicato da Microsoft nel 2007 e adottato a partire da Office 2007
  - Standard ISO/IEC 29500:2008
  - Prescrizioni *STRICT* e *TRANSITIONAL* (Documenti conformi alle prescrizioni *STRICT* sono disponibili sul mercato)
  - *Embedding* dei font/gestione dei metadati e Firma Digitale
  - Componenti del formato racchiuse in un pacchetto ZIP
- Open Document Format (ODF)
  - Standard UNI CEI ISO/IEC 26300 - ISO/IEC 26300:2006
  - Adottato come Standard documentale nativo da OpenOffice.org e da prodotti sulle principali piattaforme (Windows, Linux, Mac)
  - Larga adozione da parte delle Amministrazioni internazionali



## Altri Formati



- XML (Exstensible Markup Language)
  - Formato aperto sul quale si basano i suoi moltissimi «*Dialetti*», quali
    - XBRL – dati finanziari
    - ebXML – Commercio Elettronico
    - SOAP – scambio di messaggi tra Web Services
    - SVG – immagini vettoriali
- TXT
  - Specifica della Codifica del Carattere (Character Encoding)
- Documenti a rilevanza sanitaria
  - DICOM (immagini da strumenti diagnostici) ... non è solo un formato ma definisce anche protocolli e altro
  - HL7 - in particolare CDA2 (Clinical Document Architecture)

## Alcuni esempi



### DAX Digital Archive eXtended - Il sistema di conservazione degli archivi digitali di Regione Toscana

- Approccio flessibile che lascia aperta la possibilità di ampliare, nel tempo, i formati accettati
- Sulla base dei registri dei formati internazionali individuano i formati “consentii” e quelli “proibiti”
- Ogni ente produttore ha, però, un certo grado di discrezionalità nel decidere i formati ammissibili
- I formati ammissibili oltre quelli consigliati possono essere accettati solo previa sottomissione al sistema del/dei viewer o reader in grado di assicurarne l’esibizione (almeno su alcuni sistemi operativi).

## Alcuni esempi



### ParER - Polo archivistico regionale dell'Emilia Romagna

- Sono da preferire i formati aperti, documentati, non proprietari, diffusi, trasparenti, robusti, auto-consistenti, interoperabili, accessibili, ecc.

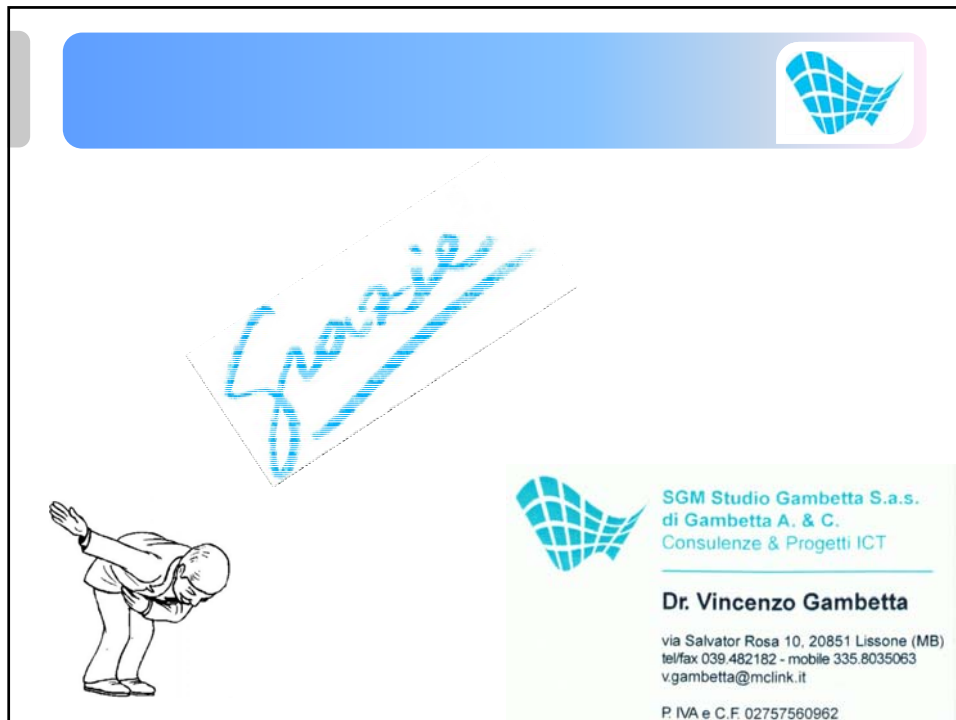
Formati	Caratteristiche	Trattamenti in fase di acquisizione	Conservazione nel lungo periodo
Accettati	Conformi ai requisiti standard previsti	Nessun trattamento	Sì
Gestiti	Non conformi ai requisiti standard ma riversabili in un formato accettato	Sono riversati in un formato accettato	Sì, nel formato riversato
Non gestiti	Non conformi ..., non riversabili ..., ma gestiti dal produttore che ne chiede comunque la conservazione	Nessun trattamento	No, è garantita solo la restituzione del documento nel formato originario
Non accettati	Non conformi ..., non riversabili ...	Rifiutati al momento del versamento	Non presenti nel sistema

## Alcuni esempi



### ParER - Polo Il sistema di Conservazione a Norma del Notariato Italiano (SCNN)

- Il sistema accetta solo formati aperti
- Documenti
  - Open Document Format (ODF) - ISO 26300
  - Portable Document Format (PDF/A) - ISO 19005
  - Rich Text Format (RTF)
  - Tutti i formati testuali (TXT, XML)
- Immagini
  - Joint Photographic Experts Group (JPEG) - ISO 10918
  - Portable Network Graphics (PNG) - ISO 15948



The business card features a blue header bar at the top right with a logo consisting of a stylized grid pattern. Below the header, the word "Grazie" is written in a large, blue, cursive font. To the left of the word is a cartoon illustration of a man in a suit bowing. To the right is a contact information box with a logo and text. The logo is a stylized grid pattern. The text in the box reads: "SGM Studio Gambetta S.a.s. di Gambetta A. & C. Consulenze & Progetti ICT", "Dr. Vincenzo Gambetta", "via Salvator Rosa 10, 20851 Lissone (MB)", "tel/fax 039.482182 - mobile 335.8035063", "vgambetta@mclink.it", and "P. IVA e C.F. 02757560962".

**SGM Studio Gambetta S.a.s.**  
di Gambetta A. & C.  
Consulenze & Progetti ICT

**Dr. Vincenzo Gambetta**

via Salvator Rosa 10, 20851 Lissone (MB)  
tel/fax 039.482182 - mobile 335.8035063  
vgambetta@mclink.it

P. IVA e C.F. 02757560962