

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

استانداردهای ذخیره و بازیابی
اطلاعات

Information Storage and Retrieval Standards



Mostafa Ghorbani

مقدمه

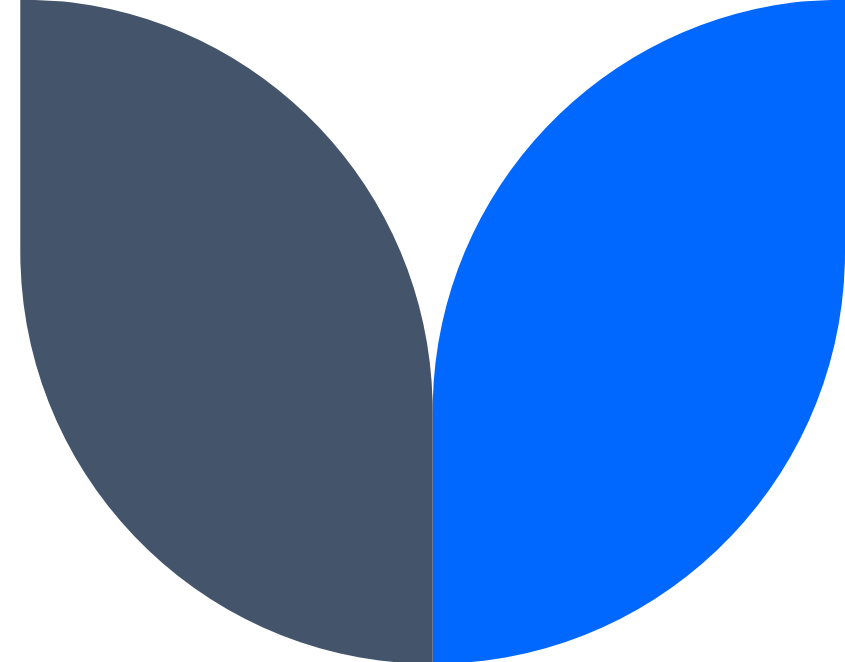
✓ استانداردها توسط طیف وسیعی از سازمان های ملی و بین المللی پشتیبانی می شوند. از جمله انجمن های حرفه ای مانند موسسه مهندسين برق و الكترونیک (IEEE)، موسسات استاندارد ملی مانند موسسه استاندارد ملی آمریکا (ANSI) یا موسسه استاندارد بریتانیا (BSI) و نهادهای بین المللی مانند سازمان بین المللی استاندارد (ISO). نکته قابل توجه برای کتابخانه ها، سازمان ملی استاندارد اطلاعات ایالات متحده (NISO) است که توسط ANSI برای تهیه استانداردهایی برای علوم کتابداری و اطلاع رسانی فعالیت می کنند.



مقدمه

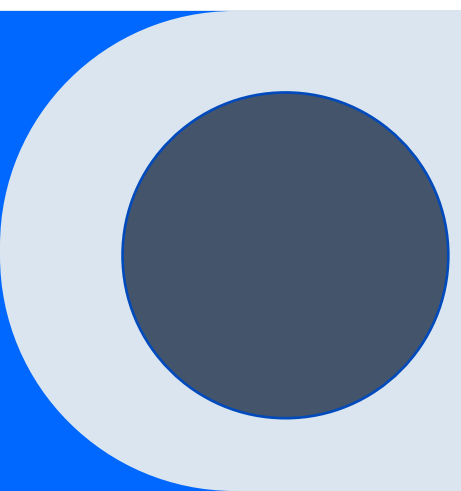
- ✓ تعدادی از موسسات و سازمان های مهم به طور فعال در توسعه و ارتقای استانداردهای مربوط به کتابخانه ها مشارکت دارند.
- ✓ فدراسیون کتابخانه های دیجیتال (DLF)، کنسرسیومی از کتابخانه ها و آژانس های مرتبط مانند آرشیو، به عنوان یکی از اهداف خود شناسایی استانداردهای مجموعه های دیجیتال و دسترسی به شبکه (<http://www.diglib.org>) را دارد.
- ✓ کتابخانه کنگره (<http://www.loc.gov>)، نقش مهمی در حفظ چندین استاندارد کلیدی مانند MARC، و توسعه MARC در یک محیط XML ایفا می کند.
- ✓ فدراسیون بین المللی انجمن ها و مؤسسات کتابخانه ای (IFLA) به منابع مربوط به انواع استانداردهای مربوطه دارد.





MARC

(Machine-Readable Cataloging)



تعريف مارک (MARC)

❖ مارک یا فهرستنویسی ماشینخوان نظام برچسبگذاری عناصر کتابشناختی برای ذخیره الکترونیک است و با استفاده از آن میتوان داده‌های موجود در یک فهرستبرگه را به‌گونه‌ای آماده ساخت که برنامه موجود در کامپیوتر بتواند همه آنها را بخواند و برابر دستور بر روی هر داده عمل کند.



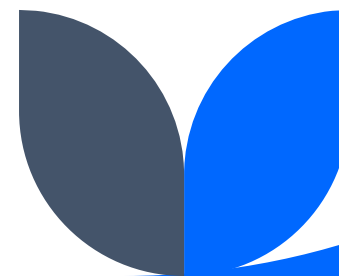
تاریخچه

✓ در اواخر دهه ۱۹۵۰ کتابخانه کنگره مطالعات مربوط به امکان استفاده از روش های خودکار را برای امور داخلی خود آغاز کرد؛ با حمایت مالی شورای منابع کتابخانه ای (CLR) بررسی عملی این امر انجام گرفت؛ این مطالعات که با علاقمندی خاصی روبرو شده بود در سال ۱۹۶۶ بصورت یک طرح درآمد و نام آن مارک نامیده شد، در این طرح ۱۶ کتابخانه شرکت داشتند؛ تا پایان سال ۱۹۶۶ رکوردهای کتابشناختی مارک بر روی نوارهای مغناطیسی بصورت هفتگی برای ۱۶ کتابخانه عضو ارسال می گردید؛ در انگلستان در سال ۱۹۶۷ اداره علمی و اطلاعاتی انگلستان به مطالعه تهیه طرح امکان یابی در مورد رکوردهای کتابشناختی ماشین خوان پرداخت؛ در سال ۱۹۷۰ به دلیل پیچیدگی طراحی فرمت مارک و همکاری نزدیک کتابخانه کنگره آمریکا و کتابخانه ملی بریتانیا مارک آمریکا و مارک انگلستان ادغام گردید؛ مارک در ابتدا بعنوان یک فرمت فهرستنویسی ماشین خوان بود اما از سال ۱۹۷۰ به عنوان فرمت تبادل داده های کتابشناختی شناخته شد و پیشرفت های زیادی نمود؛



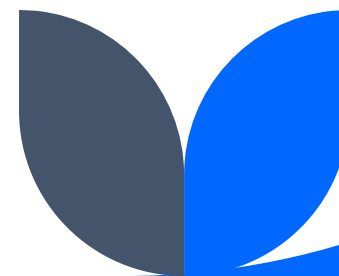
اهداف کلی

- ✓ طراحی استانداردهایی برای استفاده بهینه از کامپیوتر در جهت تبادل اطلاعات کتابشناختی بین کتابخانه ها؛
- ✓ طراحی روش هایی که به کمک آنها عملیات فهرستنویسی منابع با سرعت و کارایی بیشتری صورت گیرد؛
- ✓ امکان دستیابی کاربران به منابع اطلاعاتی در هر کجا و به هر شکل.



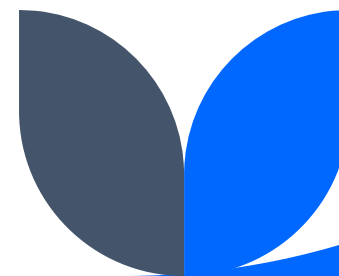
امکانات مارک

- ✓ می تواند هر رکورد را تحت یک یا چند سرعنوان خواسته شده (سرشناسه ها و شناسه های افزوده) چاپ کند و هم مدخل ها را به هر شکل دلخواه (فهرست پدیدآور، فهرست رده ای و...) مرتب کند؛
- ✓ فهرست ها را با افزایش داده های جدید به فایل مرکزی به صورت خودکار روزآمد می سازد؛
- ✓ اطلاعات مورد نیاز را از رکوردهای وارد شده بر حسب تقاضا استخراج می کند؛
- ✓ شکل نهایی فهرست یا فهرست ها را برای رسانه های مورد نیاز فراهم می کند؛
- ✓ فرمت تبادلی برای رکوردهای فهرست شده بین یک کتابخانه و سایر کتابخانه ها را فراهم می سازد.



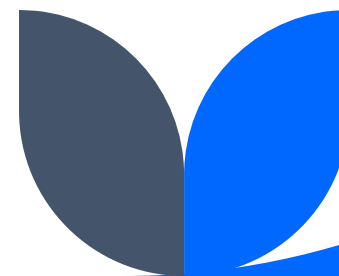
علل پیدایش مارک های ملی

- ✓ پیدایش مارک های مختلف ناشی از نقش متفاوت کتابشناسی ملی و کتابخانه ملی؛
- ✓ نبود استانداردهای فهرستنویسی بین المللی؛
- ✓ نبود نظام کنترل موضوعی بین المللی؛
- ✓ مشکلات زبانی، مجموعه حروف و کدهای متفاوت و الفبای غیر لاتینی؛



برخی از انواع مارک های ملی

مارک کانادا	INTER MARC	مارک ژاپن	JAPAN MARC
مارک فرانسه	IBER MARC	مارک کره	KOR MARC
مارک اسپانیا	AUS MARC	مارک چین	CHINES MARC
مارک استرالیا	THAI MARC	مارک کتابخانه کنگره آمریکا	LC MARC
مارک تایلند	ANNA MARC	مارک آمریکا	US MARC
مارک ایتالیا	SING MARC	مارک انگلستان	UK MARC
مارک سنگاپور	SA MARC	مارک ایران	IRAN MARC
مارک آفریقای جنوبی	INTER MARC	مارک جهانی	UNI MARC



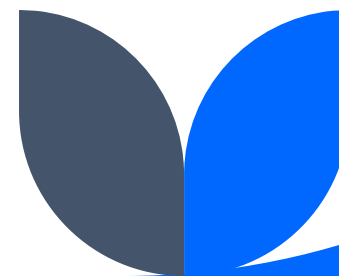
اساس طراحی مارک های مختلف

- ❖ US MARC (MARC21)
- ❖ UK MARC
- ❖ UNI MARC

گرایش های تولید مارک های ملی

✓ گرایش اول، طراحی فرمت های ملی بر مبنای یکی از دو فرمت یونی مارک و مارک امریکا؛ زیرا مارک انگلستان، به دلایلی، از این نقطه نظر کمتر مورد استفاده قرار می گیرد.

✓ گرایش دوم، روند رو به افزایش تلاش برای یکسان سازی هرچه بیشتر فرمت ها است؛ زیرا هرچه تفاوت های موجود میان فرمت های مختلف کمتر شود امکان تبادل رایانه ای داده های فهرستنویسی و کتابشناختی بیشتر می شود. از جمله این تلاش ها می توان به دیدار نمایندگان از کتابخانه کنگره امریکا و کتابخانه بریتانیا در ۱۹۹۴ اشاره کرد که با هدف هماهنگ سازی و ساده سازی دو فرمت مارک امریکا و مارک انگلستان، در راستای افزایش سازگاری میان دو فرمت، انجام گرفت.



مارك 21 (MARC21)

در سال ۱۹۹۹ USMARC و CAN/MARC برای ایجاد MARC21 ادغام شدند. این فرمت بر مبنای قوانین اداره استاندارد آمریکا ANSI با توجه به قوانین فهرستنویسی انگلو امریکن و قواعد بین المللی ISBD می باشد.



مارک جهانی (UNIMARC)

در سال ۱۹۷۷ زیر نظر ایفلا بوجود آمد. فرمتی جهانی است برای تبادل اطلاعات بین انواع مارک ها، هدف اولیه آن سهولت مبادله بین المللی داده های کتابشناختی قابل خواندن با ماشین میباشد.



ویژگی های UNIMARC

- یونی مارک از ابتدا به عنوان نوعی فرمت تبادل اطلاعات کتابشناختی میان سیستم های متفاوتی بوجود آمد و در طراحی آن متخصصان مارک های ۱۶ کشور مختلف شرکت داشتند؛
- یونی مارک فرمتی مختص یک نوع مواد کتابخانه ای نیست بلکه فرمتی یکپارچه برای همه مواد کتابخانه ای بدون در نظر گرفتن شکل ها و انواع آنها ارائه می کند؛
- از استاندارد ISBD استفاده می کند و دارای ثبات است؛
- علیرغم این که یونی مارک نوعی فرمت برای تبادل است به سادگی می توان آن را برای تبدیل به فرمت پردازش برای استفاده در یک کتابخانه یا کشور خاص تغییر داد؛
- یونی مارک برای نمایش روابط کتابشناختی میان آثار مختلف از جدیدترین روش ها استفاده می کند. این قالب با بکارگیری روش های متفاوت، می تواند روابط میان آثار کتابشناختی و همچنین اشتراک موضوع یا مولف بین این آثار را نمایش دهد.



مارک ایران IRAN MARC

✓ مارک ایران یکی از استانداردهای ذخیره و انتقال کتابشناسی در فضای کتابداری و اطلاع رسانی مدرن است که کتابخانه را قادر می سازد اطلاعات و منابع خود را در سطح ملی و بین المللی مبادله کنند؛ هدف اصلی مارک ایران تسهیل تبادل داده های کتابشناختی ماشین خوان سازمانها در سطح ملی و بین المللی است .

مارک ایران همچنین می تواند به عنوان الگویی برای طراحی فرمت های داخلی جهت ذخیره و بازیابی اطلاعات کتابشناختی ماشین خوان مورد استفاده قرار گیرد؛ مارک ایران محتواهای لازم را مانند شماره ، نشانگر ، فیلد های فرعی برای اجزای رکوردهای کتابشناختی ماشین خوان تعیین میکند و فرمت های منطقی و یکنواختی را برای ساختار فیزیکی رکورد ها ارائه میدهد؛ بنابراین مارک ایران فرمتی در برگیرنده تمامی مدارک کتاب شناختی می باشد این مدارک شامل انواع : کتاب ، مواد آرشیوی ، نقشه ، مواد دیداری شنیداری، مجلات ، پرونده های کامپیوتری و غیره می باشند.

تاریخچه مارک ایران IRAN MARC

کتابخانه ملی ایران که فهرستنویسی کامپیوتری خود را از سال ۱۳۷۳ آغاز کرده بود، برای سازگار کردن اطلاعات کتابشناختی با موازین بین المللی تحقیقات خود را آغاز نمود؛ در سال ۱۳۷۶ طرح مارک ایران به شورای پژوهشی و علمی کشور پیشنهاد شد و در همان سال مورد تصویب قرار گرفت؛ اولین جلسه مارک در سال ۱۳۷۷ با شرکت کمیته ملی مارک ایران تشکیل شد؛ با مقایسه USMARC و UNIMARC، UNIMARC مورد توافق قرار گرفت و معیار تدوین مارک ایران گردید و با استفاده از استاندارد ISO ۲۷۰۹ که خود یک استاندارد بین المللی برای تعیین ساختار رکوردهای کتابشناختی است، نوشته شد.



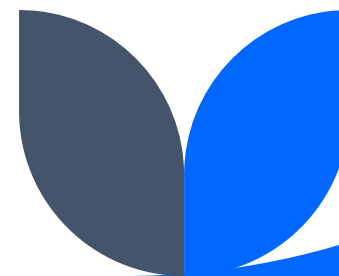
انواع فرمت مارک از نظر کاربرد و هدف

- ☐ فرمت کتابشناختی: **Bibliographic Format** این فرمت با هدف پوشش انواع مواد شامل کتاب، مواد آرشیوی و خطی، بایگانی‌های رایانه‌ای، نقشه، موسیقی، مواد بصری و ادواریها طراحی شده است.
- ☐ فرمت مستندات: **Authorities Format** این فرمت برای داده‌های کمکی به منظور استفاده کارکنانی که در ایجاد و نگهداری پیشینه‌های مستند فعال است تهیه شده است. این فرمت شامل عناصر مربوط به سرعنوان اسمی، سرعنوان نام/عنوان عنوان قراردادی، سرعنوانهای موضوعی ... است.
- ☐ فرمت موجودی: **Holdings Format** این فرمت برای عناصر موجودی کتابخانه‌های مختلف طراحی شده است.
- ☐ فرمت رده‌بندی: **Classification Format** این فرمت برای شناسایی عناصر داده‌ای در پیشینه‌های رده‌بندی طراحی شده است.



انواع فرمت مارک از نظر دامنه کاربرد

- فرمت های پردازش : **Processing Formats** که به آنها فرمت های پردازش داخلی **internal** و فرمت های کاری **Working** نیز می گویند. این فرمت برای پردازش اطلاعات در یک سیستم معین به کار می رود.
- فرمت های تبادل **Exchange Formats**: برای تبادل اطلاعات کتابشناختی بین فرمت های پردازش مختلف به کار می رود و به آنها فرمت های ارتباطات **Communications** نیز می گویند.



تعاریف مقدماتی در ذخیره اطلاعات

- ✓ کاراکتر **Character**: کوچکترین واحد اطلاعاتی را کاراکتر می گوئیم که معادل یک حرف، رقم یا نشانه (مثال x، ؟، &) می باشد، هم چنین هر کاراکتر معادل یک بایت **Byte** می باشد.
- ✓ فیلد **Field**: برای ضبط اطلاعات در مورد نمونه های مختلف یک موجودیت باید برای هر صفت خاصه یک حوزه نمایش در نظر گرفت که به این حوزه فیلد گفته می شود. بدیهی است مجموعه فیلدها یک رکورد را تشکیل میدهند.
- ✓ رکورد **Record**: یک ساختار داده ای است که مجموعه ای از فیلدها (عناصر) با نام ها و انواع داده ای مجزا می باشد. یک رکورد یا به شکل مجتمعی از عناصر و یا بطور تک تک مورد استفاده قرار می گیرد.
- ✓ فایل اطلاعاتی **Data File**: مجموعه ای از رکوردهای مرتبط به هم، تشکیل فایل اطلاعات را می دهند. به عنوان مثال مجموعه رکوردهای کتابهای فارسی، تشکیل فایل اطلاعاتی را می دهند.
- ✓ بانک اطلاعاتی **Database**: مجموعه ای از فایل های اطلاعاتی مرتبط به هم تشکیل بانک اطلاعات را می دهند. به عنوان مثال مجموعه فایل های اطلاعاتی، کتب، مقاله ها، پایان نامه ها و... بانک اطلاعات مدارک را تشکیل می دهند.

فیلدها

فیلدها **Fields**: زنجیره ای از کاراکترها هستند که بوسیله یک شماره شناسایی می شوند (۰۰۱ تا ۹۹۹) که اصطلاحاً به آن شماره “تگ” **Tag** میگویند؛

• فیلدهای فرعی **Subfields**: واحد مشخصی از اطلاعات داخل یک فیلد است در هر فیلد می توان یک یا چند فیلد فرعی داشت .



بلوک های اصلی مارک ایران

- 0 بلوک شناسایی Identification Block
- 1 بلوک اطلاعات کد شد Coded Information Block
- 2 بلوک اطلاعات توصیفی Descriptive information Block
- 3 بلوک یادداشت ها Note Block
- 4 بلوک شناسه رابط Linkage Entry Block
- 5 بلوک عنوانهای مرتبط Related title Block
- 6 بلوک تجزیه و تحلیل موضوعی Subject Analysis Block
- 7 بلوک مسوولیت معنوی اثر Intellectual Responsibility Block
- 8 بلوک ویژه کاربرد بین المللی International Use Block
- 9 بلوک ویژه استفاده کتابخانه های ایران National Use Block





Dublin Core

تاریخچه Dublin Core

✓ Dublin Core یک استاندارد فراداده است که شامل مجموعه ای از عناصر داده است که می تواند برای توصیف اسناد دیجیتال استفاده شود. ابتکار فراداده هسته دوبلین (DCMI) مسئول فرمول بندی هسته دوبلین است. DCMI پروژه ای از انجمن علوم و فناوری اطلاعات (ASIS&T)، و یک سازمان غیرانتفاعی است. دوبلین کور در یک سری کارگاه های آموزشی با حضور نمایندگان از سراسر جهان آغاز شد که از سال ۱۹۹۵ شروع شد، که به تعریف و توسعه الزامات ابرداده برای استفاده در اینترنت و همچنین در برنامه های محلی، به طوری که به کتابخانه های دیجیتال دیگران نیز بتوانند به آن دست یابند. نام آن از دوبلین، اوهایو، جایی که اولین کارگاه در آنجا برگزار شد، می آید. Dublin Core، به عنوان یک استاندارد توسط NISO در سپتامبر ۲۰۰۱ و توسط ISO در فوریه ۲۰۰۳ پذیرفته شد.



Dublin Core

- ✓ **Dublin Core** هیچ طرح کنترل واژگان خاصی را برای نشان دادن محتوای موضوعی اعمال نمی کند.
- ✓ استاندارد **Dublin Core** در ابتدا شامل دو سطح بود:
- ✓ **Simple Dublin Core** شامل ۱۵ عنصر است.
- ✓ **Qualified Dublin** شامل سه عنصر اضافی (مخاطب، منشأ و حق مؤلف) برای ارائه ویژگی و جزئیات بیشتر است.



Dublin Core

- ✓ ۱۵ عنصر اساسی Dublin Core که همه اسناد الکترونیکی باید برای قابلیت همکاری داشته باشند:
- ✓ عنوان: نامی که به سند داده شده است
- ✓ سازنده: نهادی که مسئولیت اصلی ایجاد محتوا را بر عهده دارد.
- ✓ موضوع: موضوع محتوا، معمولاً کلمات کلیدی یا کدهای طبقه بندی که از یک واژگان کنترل شده یا طرح طبقه بندی رسمی گرفته شده است.
- ✓ توضیحات: گزارشی از محتوا، مانند چکیده یا فهرست مطالب است:
- ✓ ناشر: نهادی که مسئول در دسترس قرار دادن منبع است.
- ✓ مشارکت کننده: نهادی که مسئول مشارکت در محتوا است.
- ✓ تاریخ: تاریخ یک رویداد در چرخه عمر سند است.
- ✓ نوع: ماهیت محتوا.
- ✓ قالب: جلوه فیزیکی یا دیجیتالی سند.
- ✓ شناسه: یک ارجاع بدون ابهام به سند در یک زمینه خاص، مانند یک منبع یاب یکنواخت (URL یا ISBN)
- ✓ منبع: ارجاع به منبعی که سند از آن مشتق شده است
- ✓ زبان: زبان محتوای سند
- ✓ رابطه: ارجاع به یک سند مرتبط
- ✓ پوشش: میزان یا دامنه محتوا، معمولاً مکانی، زمانی یا قانونی
- ✓ حقوق: اطلاعاتی در مورد حقوقی که در سند و روی آن نگهداری می شود.

Dublin Core Elements

Rights

Contributor

Creator

Subject

Coverage

Title

Publisher

Identifier

Description

Type

Date

Source

Relation

Format

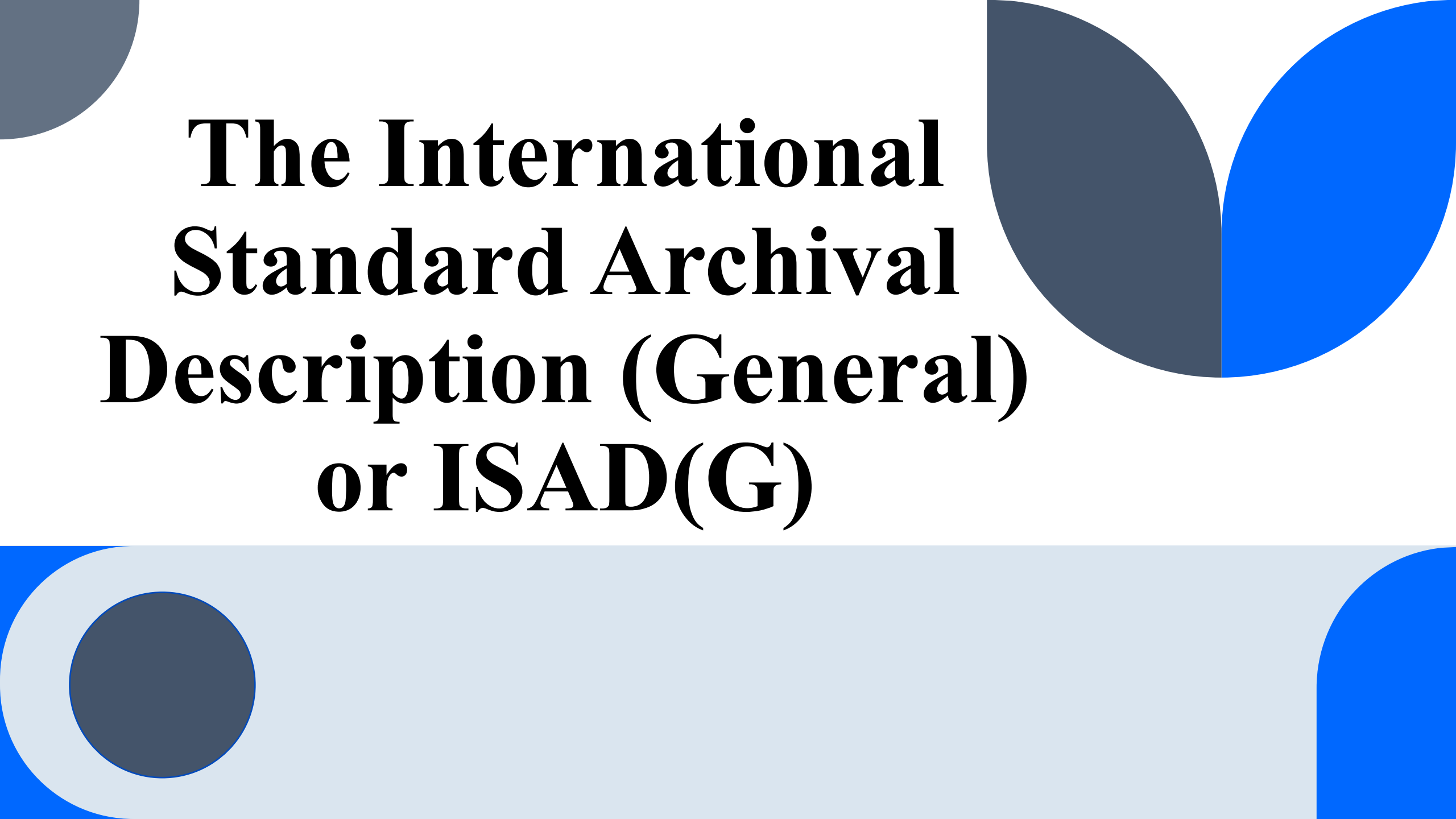
Language

Dublin Core in HTML

```
<html>
  <head>
    <meta name="DC.title" content="The Card Turner" />
    <meta name="DC.creator" content="Louis Sachar" />
  </head>
  <body>
    ...
  </body>
</html>
```

Dublin Core in XML

```
<metadata>  
  <dc:title>The Card Turner<dc:title>  
  <dc:creator>Louis Sachar<dc:creator>  
</metadata>
```



The International Standard Archival Description (General) or ISAD(G)

شرح آرشیو استاندارد بین المللی (عمومی) یا ISAD(G)

- ✓ توسط کمیته ای از شورای بین المللی آرشیو ایجاد شده است، راهنمایی های کلی را در مورد تهیه توصیفات آرشیوی به منظور شناسایی و توضیح زمینه و محتوای مواد آرشیوی و ارتقاء دسترسی به آنها ارائه می دهد.
- ✓ این استاندارد که برای اولین بار در سال ۱۹۹۶ ظاهر شد، شامل ۲۶ عنصر است که می توان آنها را ترکیب کرد تا توصیف یک موجودیت آرشیوی را ایجاد کند.



The background features abstract geometric shapes in blue and grey. In the top right, there are two overlapping semi-circles, one grey and one blue. In the bottom left, there is a grey circle within a larger blue circle. The bottom half of the image has a light blue background with a dark blue rounded rectangle in the bottom right corner.

Encoded Archival Description (EAD)

شرح بایگانی رمزگذاری شده (EAD)

✓ EAD به طور گسترده توسط بایگانی‌ها، موزه‌ها و کتابخانه‌های نسخ خطی به عنوان استاندارد برای رمزگذاری ابزارهای کمکی، راهنماها، فهرست‌های دستی یا کاتالوگ‌ها استفاده می‌شود. به طور مشترک توسط انجمن آرشیوداران آمریکا و کتابخانه کنگره اداره می‌شود. هدف آن فعال کردن آرشیوها و کتابخانه‌ها برای به اشتراک گذاشتن اطلاعات در مورد رکوردها و مجموعه‌های مرتبط اما متفاوت است.



شرح بایگانی رمزگذاری شده (EAD)

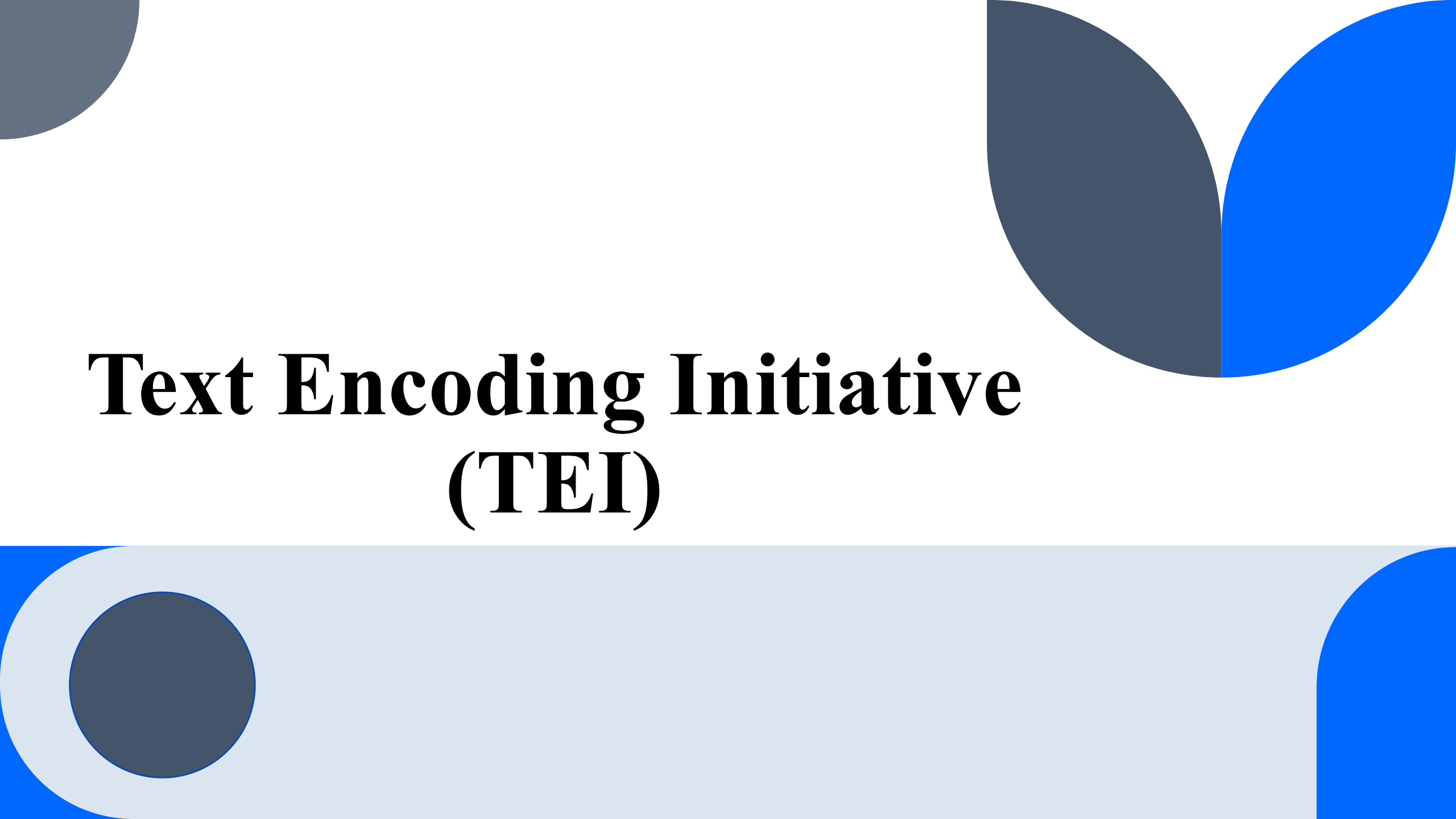
- ✓ EAD در سال ۱۹۹۳ در دانشگاه کالیفرنیا به عنوان سرپرست پروژه توسط Daniel V. Pitti آغاز شد.
- ✓ EAD به طور رسمی به عنوان SGML DTD مشخص می شود، اما اکنون به عنوان طرحواره XML نیز مشخص می شود.
- ✓ نسخه فعلی EAD 2002 نام دارد. این نسخه توسط کتابخانه کنگره نگهداری می شود.



شرح بایگانی رمزگذاری شده (EAD)

- ✓ بایگانی‌ها را قادر می‌سازد تا اطلاعات مربوط به مواد مرتبط اما متفاوت و مطالب پراکنده را به اشتراک بگذارند.
- ✓ توضیحات منسجم و قابل فهمی را در مخازن مختلف ارائه می‌دهد.
- ✓ دسترسی جهانی به مجموعه‌های توزیع شده را ارائه می‌دهد.





Text Encoding Initiative (TEI)

ایتکار رمزگذاری متن (TEI)

- ✓ TEI یک استاندارد بین‌المللی و بین‌رشته‌ای برای نشان دادن آن دسته از ویژگی‌های یک متن است که باید به صراحت شناسایی شوند، تا پردازش متن کامپیوتری و اشتراک‌گذاری متون الکترونیکی تسهیل شود.
- ✓ شامل مجموعه‌ای از تگ‌ها است که می‌توان آن را در متن وارد کرد تا ساختار آن را مشخص کند.



ابتکار رمزگذاری متن (TEI)

- ✓ Text Encoding Initiative در سال ۱۹۸۷ تاسیس شد.
- ✓ رهنمودهایی برای رمزگذاری متون قابل خواندن توسط ماشین در علوم انسانی و اجتماعی است.
- ✓ شامل راه های استاندارد شده برای علامت گذاری معنایی یک سند است.





با تشکر از حسن توجه شما ...