



FINANCIALTM
DATA EXCHANGE

Tax Document PDF
Embedded Data

Version 2024.0
June 2024



Legal Notice

Financial Data Exchange, LLC (FDX) is a standards body and provides these materials for general use among industry stakeholders. Many of the terms, however, are subject to additional interpretations under prevailing laws, industry norms, and/or governmental regulations. While referencing certain laws that may be applicable, readers, users, members, or any other parties should seek legal advice of counsel relating to their particular practices and applicable laws in the jurisdictions where they do business. See FDX's complete Legal Disclaimer located at <https://www.financialdataexchange.org> for other applicable disclaimers. The information provided herein is for educational purposes only and is not intended to be a guide for any specific company. Each company should consult with its own legal, IT, data security, financial, tax, and other advisors before implementing any programs described herein. References to the U.S. market, U.S. laws, and the like, will require certain modification or analysis to confirm applicability in other jurisdictions.

References to other sites, sample code, and resources provided by third parties are provided for your convenience only and shall not be deemed an endorsement nor recommendation by FDX. Each party is subject to complying with all applicable license agreements and FDX grants no such rights. Any use of these open source tools and other materials referenced here are governed by each tool's own license. Refer to and understand those licenses before adopting and using in your environment at your own risk and obligation for compliance thereto. We have no control over the contents of those sites or resources and accept no responsibility for them or for any loss or damage that may arise from your use of them. If you decide to access any of the third-party websites linked herein, software, toolsets, sample code, and/or open source materials, you do so entirely at your own risk and subject to the terms and conditions of use for such websites and third-party rights. FDX asserts no ownership, copyright, or other claims to any third-party materials, software, or other tools referenced or mentioned herein.

If you are not a licensed user of the FDX API, the use of the FDX API is subject to your acceptance of the [FDX API License Agreement](#), as amended. The FDX API License Agreement, as well as your acceptance of the FDX Terms of Use, and FDX Privacy Policy must first be accepted by contacting FDX at <https://www.financialdataexchange.org> and selecting "Get Started."

Revision History

Document Version	Notes	Date
2024.0	Updates to keys, banners, and examples.	June 2024
2021.0	Initial Document Release	October 2021

Contents

INTRODUCTION	4
TAXPAYER BENEFITS	4
TAX AND FINANCIAL INDUSTRY BENEFITS	4
AUDIENCE	5
TAX DOCUMENT ISSUERS	5
TAX INFORMATIONAL STATEMENTS	6
SHARING TAX DATA THROUGH PDF FILES WITH EMBEDDED FDX JSON	7
PDF WORKFLOW	7
DOCUMENT ISSUER IMPLEMENTATION STEPS	8
IDENTIFICATION BANNER	9
FINANCIAL DATA EXCHANGE TAX API	9
TAX SOFTWARE IMPLEMENTATION STEPS	9
EXAMPLE: EMBEDDED PDF DATA	10
<i>FDX Tax Data for IRS Form 1098</i>	10
EMBEDDED TAX DATA AS PDF CUSTOM PROPERTIES	13
APPENDIX 1: TECHNICAL INFORMATION FOR PDF FILE GENERATION USING JAVA	14
APPENDIX 2: TECHNICAL INFORMATION FOR READING PDF EMBEDDED DATA USING JAVA	15

Introduction

One of the goals of the Financial Data Exchange organization is to define standards and encourage practices that improve the experience of consumers in the management of their financial data. Each year consumers must manage the receipt of annual tax documents and use the data within in the preparation of their income tax returns.

This document defines a mechanism to support tax data management in taxpayers' annual tax documents that are provided in Portable Document Format (PDF) files.

Taxpayer benefits

Taxpayer benefits include:

- Eliminating data entry errors in the income tax return preparation process
- Reducing time spent in tax preparation
- Potential to expand the types of documents that can be imported into tax software
- Potential to expand the number of tax software applications supporting data import

Tax and financial industry benefits

Tax and financial industry benefits include:

- Reduced programming costs of simpler systems
- Eliminating the use of Optical Character Recognition (OCR) technology and its potential errors
- Eliminating the use of custom PDF text-parsing software
- A simpler mechanism for tax document issuers to integrate with tax software:
 - Taxpayer consent is implicit through their file upload. No need to capture, store and authenticate consent
 - Technology connections to industry partners not needed
 - Full-fledged API services consuming or providing data also not needed

Audience

This document defines a simple mechanism which the tax industry can use to support taxpayers' tax data management. The intended audiences are:

- Tax Document Issuers: Organizations which have IRS-mandated requirements to deliver tax documents to taxpayers, or offer other financial information statements for taxpayer convenience
- Tax Software Companies: Builders of applications which use data from tax documents or informational statements to prepare tax returns

The IRS hosts information on the various tax information forms which must be sent to taxpayers at <https://www.irs.gov/instructions/i1099gi>.

Tax Document Issuers

Examples of organizations and some of the tax documents which they issue:

- Banks – 1099-INT, 1099-DIV, 1099-OID, 1099-B
- Lenders – 1098, 1098-E, 1099-A, 1099-C
- Brokerage houses – 1099-INT, 1099-DIV, 1099-OID, 1099-B, 1097-BTC
- Insurance companies – 1095-A, 1095-B, 1095-C
- Corporations – 2439, 3921, 3922, 1099-DIV, 1099-MISC, 1099-NEC
- Employers – W-2's, 1099-MISC, 1099-NEC
- Government agencies – 1099-G, 1099-Q
- Payroll providers – W-2's
- Universities – 1098-T
- Partnerships – 1065 K-1
- Escrow companies – 1099-S
- Trusts – 1041 K-1
- Cooperatives – 1099-PATR
- Payment card providers – 1099-K
- Pension plans – 1099-R
- IRA trustees – 1099-R, 5498

Tax Informational Statements

The FDX U.S. Tax specification and this Embedded Data in PDF Tax Documents specification also provide support for informational statements of income and expense for these additional activities:

- Farms – Farm Income Statement (for IRS Schedule F), Farm Rental Statement (for IRS Form 4835)
- Real Estate management companies – Rental Income Statement (for IRS Schedule E)
- Oil & Gas industry – Royalty Income Statement (for IRS Schedule E)
- Small Businesses – Business Income Statement (for IRS Schedule C)
- Crypto traders – Cryptocurrency Tax Statement (for IRS Form 8949)

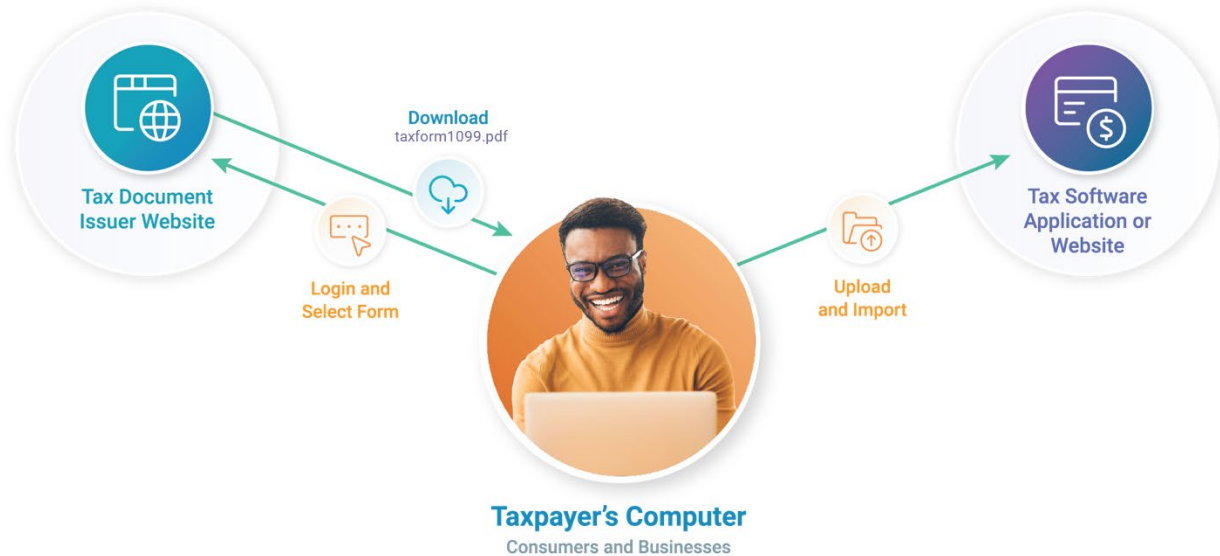
Sharing Tax Data through PDF Files with Embedded FDX JSON

PDF files provide for embedding textual information in what are known as custom properties. Custom properties are key-value pairs with arbitrary keys. Document issuers will include the tax document data serialized to JSON format in a custom property with a key of **fdxJson**.

PDF Workflow

A taxpayer logs in to the website hosting their tax document(s) and downloads the PDF file(s) provided. The taxpayer then uploads the PDF file(s) to the tax software application.

This diagram illustrates the workflow.



©2021  FINANCIAL
DATA EXCHANGE

Document Issuer Implementation Steps

To implement this feature, developers programmatically add the below custom property keys and their related values to the PDF.

Key 1: **fdxVersion**

A string containing the version number of the FDX Specification used to create the FDX JSON content referenced below, for example "V5.3.3" or "V6.0.0". The FDX versions are enumerated in the `FdxVersion` type defined in the `fdxapi.components.yaml` file of the FDX API. This field informs tax software which API version was used to serialize the data in `fdxJson`.

Key 2: **fdxSoftwareId**

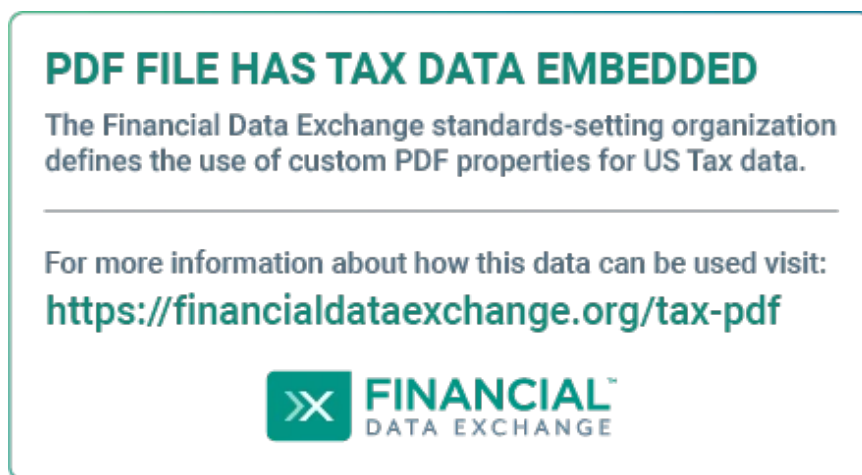
The ID assigned to the software which generated the PDF or just the embedded data. A company can register for a software ID with the Financial Data Exchange. This is expected to be a unique human-readable name. Tax software can use this to identify and deal with any peculiarities in the data generation done by the software with that ID.

Key 3: **fdxJson**

The PDF generation collects the document data and converts it to an instance of the FDX Tax Data API's `TaxDataList` entity (through V5.3.3) or `TaxStatement` entity (beginning V6.0.0). The instance is serialized to JSON and placed in the **fdxJson** custom property. See the last section for a complete example.

Identification Banner

Producers of tax document PDFs containing embedded tax data are encouraged to identify and promote that functionality by displaying on the PDF an FDX-defined image banner (or the corresponding text). Banner images are available at <https://financialdataexchange.org/tax-pdf> in bitmap .png, vector graphic .ai and .pdf file formats to support your technology choice.



Financial Data Exchange Tax API

The full FDX specifications and JSON schemas, including this document, are available from [here](#). FDX content is free to non-member organizations by registering [here](#).

Tax Software Implementation Steps

Importing and using tax data from a taxpayer's PDF with FDX JSON comprises these steps:

1. The taxpayer signs into a tax software application
2. The tax software application accepts upload of the taxpayer's PDF file
3. The tax software application programmatically opens the PDF file and reads the value of the **fdxVersion**, **fdxSoftwareId**, and **fdxJson** custom properties
4. The application deserializes the **fdxJson** custom property value and imports the FDX JSON data securely and accurately to prepare the taxpayer's income tax return (no OCR errors introduced)

Example: Embedded PDF Data

IRS Form 1098 Example

The following is an example (with mock data) of an IRS Form 1098 that could be delivered to a taxpayer as a digital PDF file, showing how its Embedded PDF Data banner could appear in whitespace near the mailing label.

☐ CORRECTED (if checked)

RECIPIENT'S/LENDER'S name, street address, city or town, state or province, country, ZIP or foreign postal code, and telephone no. Financial Intelligence Associates 12022 Sundown Valley Dr Suite 230 Reston, VA 20191 888-555-1212		*Caution: The amount shown may not be fully deductible by you. Limits based on the loan amount and the cost and value of the secured property may apply. Also, you may only deduct interest to the extent it was incurred by you, actually paid by you, and not reimbursed by another person.		OMB No. 1545-1380 Form 1098 (Rev. January 2022) For calendar year 2023	Mortgage Interest Statement Copy B For Payer/ Borrower The information in boxes 1 through 9 and 11 is important tax information and is being furnished to the IRS. If you are required to file a return, a negligence penalty or other sanction may be imposed on you if the IRS determines that an underpayment of tax results because you overstated a deduction for this mortgage interest or for these points, reported in boxes 1 and 6; or because you didn't report the refund of interest (box 4); or because you claimed a nondeductible item.
RECIPIENT'S/LENDER'S TIN 12-3456789		PAYER'S/BORROWER'S TIN xxx-xx-1234		1 Mortgage interest received from payer(s)/borrower(s) \$ 1008.00	
PAYER'S/BORROWER'S name Kris Q Public		2 Outstanding mortgage principal \$ 200900.00	3 Mortgage origination date 03/10/2022	4 Refund of overpaid interest \$ 4011.00	
Street address (including apt. no.) 1 Main St		5 Mortgage insurance premiums \$ 5012.00	6 Points paid on purchase of principal residence \$ 6013.00	7 <input checked="" type="checkbox"/> If address of property securing mortgage is the same as PAYER'S/BORROWER'S address, the box is checked, or the address or description is entered in box 8.	
City or town, state or province, country, and ZIP or foreign postal code Melrose, NY 12121		8 Address or description of property securing mortgage			
9 Number of properties securing the mortgage 9	10 Other 10. Property tax: \$10,017.00		11 Mortgage acquisition date 11/15/2022		
Account number (see instructions) 111-23456					

Form **1098** (Rev. 1-2022) (Keep for your records) www.irs.gov/Form1098 Department of the Treasury - Internal Revenue Service

PDF FILE HAS TAX DATA EMBEDDED

The Financial Data Exchange standards-setting organization defines the use of custom PDF properties for US Tax data.

For more information about how this data can be used visit:
<https://financialdataexchange.org/tax-pdf>



FDX Tax Data for IRS Form 1098

The following is the FDX-defined tax data format for the three custom attributes using the mock data shown on above form image.

```
"fdxVersion": "V6.0.0"
```

```
"fdxSoftwareId": "FdxBankSoftware-2024"
```

```

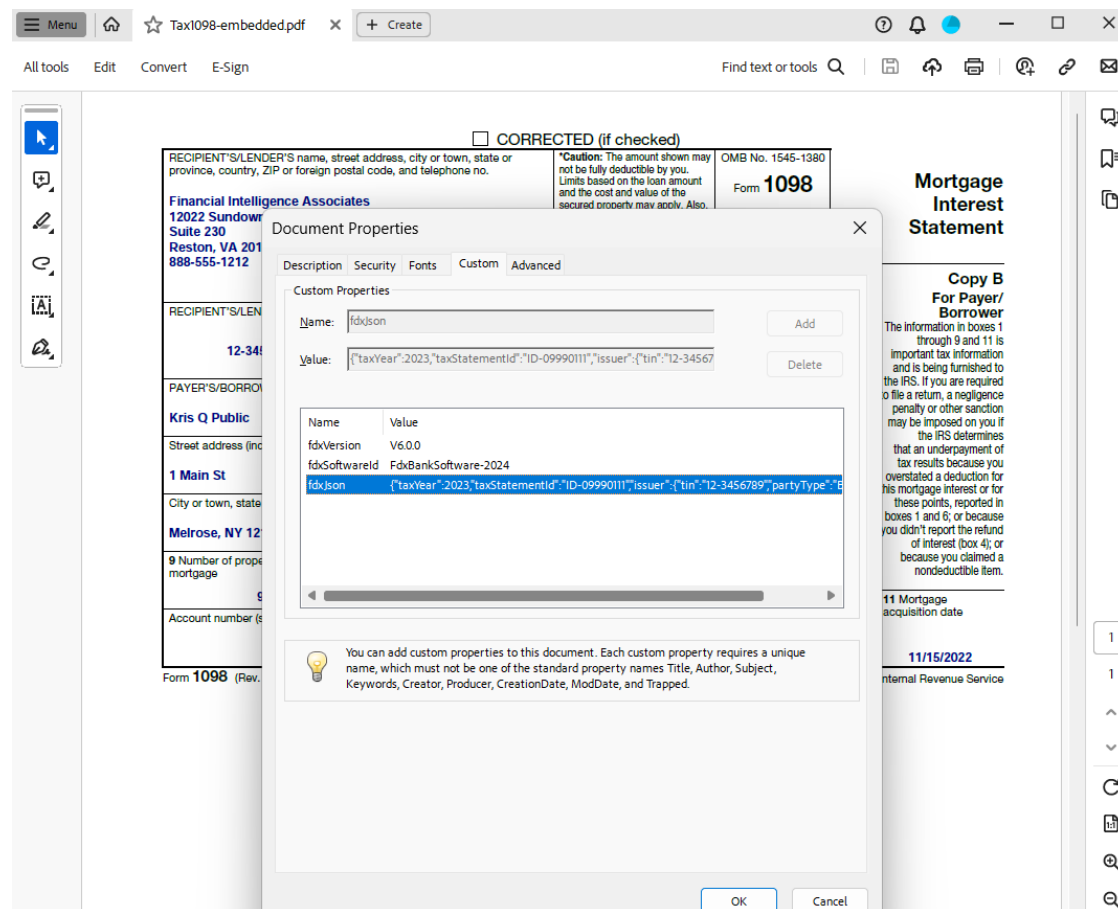
"fdxJson": {
  "taxYear": 2023,
  "taxStatementId": "ID-09990111",
  "issuer": {
    "tin": "12-3456789",
    "partyType": "BUSINESS",
    "businessName": {
      "name1": "Financial Intelligence Associates"
    },
    "address": {
      "line1": "12022 Sundown Valley Dr",
      "line2": "Suite 230",
      "city": "Reston",
      "region": "VA",
      "postalCode": "20191",
      "country": "US"
    },
    "phone": {
      "number": "8885551212"
    }
  },
  "recipient": {
    "tin": "xxx-xx-1234",
    "partyType": "INDIVIDUAL",
    "individualName": {
      "first": "Kris",
      "middle": "Q",
      "last": "Public"
    },
    "address": {
      "line1": "1 Main St",
      "city": "Melrose",
      "region": "NY",
      "postalCode": "12121",
      "country": "US"
    }
  },
  "forms": [
    {
      "tax1098": {
        "taxYear": 2023,
        "taxFormId": "ID-09990111-1098",

```

```
    "taxFormDate": "2024-02-15",
    "taxFormType": "Tax1098",
    "mortgagedProperties": 9,
    "otherInformation": "10. Property tax: $10,017.00",
    "accountNumber": "111-23456",
    "mortgageInterest": 1008.0,
    "outstandingPrincipal": 200900.0,
    "originationDate": "2022-03-10",
    "overpaidRefund": 4011.0,
    "mortgageInsurance": 5012.0,
    "pointsPaid": 6013.0,
    "isPropertyAddressSameAsBorrowerAddress": true,
    "acquisitionDate": "2022-11-15",
    "propertyTax": 10017.0
  }
}
]
```

Embedded Tax Data as PDF Custom Properties

The recipient of a digital PDF file with tax data embedded as custom properties can themselves view those properties using Acrobat® Reader® on their computer. When viewing the file within Acrobat simply Select the menu option **File** → **Document Properties** and on the resulting pop-up Document Properties window select the **Custom** tab.



This can be performed on this very PDF file to see the example custom properties shown above.

Appendix 1: Technical information for PDF file generation using Java

The following Java class illustrates how to embed custom properties in a PDF file using the Apache PDFBox library. While there are other PDF libraries available, this code example is likely similar to what is needed for all of them.

```
import org.apache.pdfbox.pdmodel.PDDocument;
import org.apache.pdfbox.pdmodel.PDDocumentInformation;
import java.io.File;
import java.io.IOException;

public class PdfFdxJsonInjector {

    public void inject( String pdfInFilePath,
                        String pdfOutFilePath,
                        String fdxVersion,
                        String fdxSoftwareId,
                        String fdxJson ) throws IOException {

        try(
            PDDocument document = PDDocument.load( new File( pdfInFilePath ) );
        ) {

            // Inject custom properties
            PDDocumentInformation documentInformation =
                new PDDocumentInformation( );

            documentInformation.setCustomMetadataValue(
                "fdxJson",
                fdxJson
            );

            documentInformation.setCustomMetadataValue(
                "fdxVersion",
                fdxVersion
            );

            documentInformation.setCustomMetadataValue(
                "fdxSoftwareId",
                fdxSoftwareId
            );

            document.setDocumentInformation( documentInformation );

            document.save( new File( pdfOutFilePath ) );

        } catch ( Exception e ) {
            System.out.println( e.getMessage( ) );
        }
    }
}
```

Appendix 2: Technical information for reading PDF embedded data using Java

The below Java class illustrates how to read custom properties from a PDF file using the Apache PDFBox library. While there are other PDF libraries available, this code example is likely similar to what is needed for all of them.

```
import org.apache.pdfbox.pdmodel.PDDocument;
import org.apache.pdfbox.pdmodel.PDDocumentInformation;

import java.io.File;

public class PdfFdxJsonReader {

    private String fdxJson;
    private String fdxVersion;
    private String fdxSoftwareId;

    public String getFdxJson( ) {
        return fdxJson;
    }

    public String getFdxVersion( ) {
        return fdxVersion;
    }

    public String getFdxSoftwareId( ) {
        return fdxSoftwareId;
    }

    public void read(
        String pdfInFilePath
    ) {
        try(
            PDDocument document = PDDocument.load( new File( pdfInFilePath) );
        ) {

            PDDocumentInformation info
                = document.getDocumentInformation( );

            if ( info == null ) {
                return;
            }

            this.fdxJson = info.getCustomMetadataValue( "fdxJson" );
            this.fdxVersion = info.getCustomMetadataValue( "fdxVersion" );
            this.fdxSoftwareId = info.getCustomMetadataValue( "fdxSoftwareId" );

        } catch ( Exception e ) {
            System.out.println( e.getMessage( ) );
        }
    }
}
```