

MAY 2026

The Future of Tax Filing

Part 2: Building Direct File: Policy and Strategy

Chapter 2: Direct File and data import / pre-population

Gabriel Zucker, Chris Given

Contents

[2.1 Conceptual framework and ‘return-free filing’](#)

[2.1.1 The importance of pre-population and the goals of Direct File](#)

[2.1.2 Return-free filing as the endpoint of Direct File, not a separate project](#)

[2.1.3 Pre-population, pre-fill, data import, simplified filing, and return-free filing – defining our terms](#)

[2.2 Data sources and limitations](#)

[2.2.1 Third-party data, or information returns \(W-2s, 1099s, 1095s, 5498s\)](#)

[2.2.1.1 W-2](#)

[2.2.1.2 Other information returns](#)

[2.2.1.3 Timing of information returns](#)

[2.2.1.4 Errors in information returns](#)

[2.2.2 Data from taxpayer account or identity proofing](#)

[2.2.3 Data from other current-year returns](#)

[2.2.4 Data from prior-year returns](#)

[2.2.5 Data from SSA](#)

[2.2.6 Data held by the state tax department](#)

[2.2.7 Technical issues with non-primary taxpayers](#)

[2.3 Limitations of Direct File functionality built to date, and expansion considerations](#)

[2.4 Next steps and recommendations](#)

[2.4.1 Expanding pre-population to include more data items](#)

[2.4.2 Pre-population, Direct File architecture, and the taxpayer experience](#)

[2.4.3 Pre-population and state data](#)

[2.4.4 Information return availability and reliability](#)

Summary

- Expanding *pre-population* — the use of IRS data to streamline the tax filing process — has the potential to make tax filing quick, easy, and painless; it would end tax filing as we know it. (2.1.1)
- Within the confines of the existing U.S. tax system, though, we cannot eradicate tax returns altogether. Attempting to do so would, in fact, likely hurt the taxpayers we are trying to help. The most automated tax system we can achieve is a Direct File-like product with maximum pre-population, and reformers should keep energy on that track, rather than attempting to build a parallel and more-automated track. (2.1.2)
- Maximizing pre-population in Direct File will require technical improvements in the timing and processing of third-party information returns. (2.2.1)
- It will also require using data from a number of other data sources (2.2.2-2.2.5), though perhaps not from state tax departments (2.2.6).
- Next steps and recommendations:
 - Direct File should continue to expand pre-population to include more data sources and data items, remembering that delivering pre-population is an iterative process. (2.4.1)
 - Direct File may have to reimagine the connection of pre-population to the taxpayer experience, to achieve the full promise of the functionality. (2.4.2)
 - Direct File should work with state partners to explore the pre-population of state tax department data, though it is possible the effort required will not be worth it. (2.4.3)
 - A variety of improvements will be needed to ensure the sufficiency of third-party information return data, including retrieving and storing W-2 Boxes 12-20 data, and changing reporting deadlines for various Forms 1099 (which will likely require statutory changes to 26 USC 6071). (2.4.4)

This chapter explores the past and potential future of *data import* or *pre-population* in Direct File — the use of IRS data to transform the filing process.

Section 2.1 lays out the importance of this topic to the project of Direct File, explores the conceptual limits of what is possible in the U.S. tax system, and defines terms. Section 2.2 explores pre-population data elements source by source, exploring specific challenges and deficiencies in the available data. Section 2.3 provides an overview of relevant Direct File functionality built to date. Section 2.4 recommends next steps.

2.1 Conceptual framework and ‘return-free filing’

2.1.1 The importance of pre-population and the goals of Direct File

Of Direct File’s various benefits, the element that has arguably most inspired filing reform advocates is its ability to streamline the filing process by using data the IRS already has.

In part, pre-population’s appeal is rooted in simple common sense. Taxpayers know that their data is reported to the IRS, so much so that if they make a mistake on their return, the IRS very well may notice, correct it, and even impose consequences. If the IRS already has the answers to this particular test, taxpayers say, why is it making us guess?¹ **Using IRS data to streamline return filing would remove the exasperation from tax filing and alter taxpayers’ impression that the IRS is trying to make life harder instead of easier.**

“Using IRS data to streamline or automate tax returns is therefore critical to closing the tax benefits gap.”

But the appeal is also motivated by the persistent tax benefits access gap. While some commonly-cited figures may be overestimates (see Chapters [12](#) on dependent rules and [14](#) on non-filers), there are millions of households a year who leave billions of dollars in tax benefits on the table because they do not file returns. The [unambiguous lesson of “simplified filing” via GetCTC in 2021 and 2022](#) is that a filing product that does not require taxpayers to transcribe their tax documents or answer esoteric tax questions is the most powerful way

to break down barriers for this population. **Using IRS data to streamline or automate tax returns is therefore critical to closing the tax benefits gap, unlocking billions in unclaimed benefits for those Americans who need them most.**

Direct File began to implement pre-population via prior-year-AGI import in 2024 and added several more pieces of pre-population in 2025, including W-2 import. As pre-population expanded, it would have come to include more data sources, and the filing experience would have been rebuilt around that data. This means not simply piping text into some fields and skipping over the occasional page, but actually designing an experience that draws powerful conclusions from the data that exists, and allows many taxpayers to move through the filing process incredibly quickly.

¹ In the words of the May 2023 Direct File Report to Congress: “Some interviewed taxpayers know that tax forms such as the W-2 and 1099s are sent to both them and IRS. Some expressed a preference for this data to be used to streamline the experience of an IRS-provided tool. After using the functioning internal Direct File prototype without this capability in a user research interview, one interviewed taxpayer said, ‘IRS already knows your tax information. So why wouldn’t I be able to login, put in, say, my [SSN], and then half this information is already filled, and then I just need to put in corrections, you know?’”

We believe the impact for the perception and reality of tax filing would be immense. Gone would be the days when even relatively simple returns can take hours to complete, when tracking down and collating W-2s and 1099s is a weeklong chore, where even conscientious taxpayers live with the nagging suspicion they forgot something and may get in trouble. Gone would be the days of manually transcribing esoteric minutiae and wondering which details are really relevant to feed into the black box that is filing software. **Tax filing would be quick, painless, and easy. We would indeed have ended tax filing as we know it.**

2.1.2 Return-free filing as the endpoint of Direct File, not a separate project

At the same time, it is critical to understand the constraints of automation and pre-population — due not simply to imperfect IRS systems, but to the nature of the U.S. tax law. **In short, absent a wholesale reimagining of federal and state/local tax law in the United States, eradicating tax returns altogether, or creating a system where filing is generally as simple as answering a text message (a common refrain in some policy circles) is basically impossible.^{2,3} But we can — and should — use Direct File to iteratively make filing easier and easier, until it has ended tax filing as we know it. Insofar as advocates or policymakers want to achieve “return-free filing,” the way to do so is to keep pushing Direct File farther.**

Advocates often point to precedents in other countries, where tax filing effectively does not exist. But, unlike those foreign tax codes, the U.S. tax system requires taxpayers to provide information that tax agencies do not have and are in no position to get. The IRS in particular and the U.S. government in general do not categorically know about your child care expenses,⁴ changes in your family structure or marital status, where you lived for how many months, and much more. Even if the IRS used data from every government agency to prepare the most precise possible return for you, your calculated refund could easily be far too large or too small. Here are just a few examples of situations where automatically issuing a return without taxpayer involvement could cause taxpayers to *overpay* their taxes:

- Last tax year, you took in your sister’s son, living with him for most of the year and becoming his primary caregiver. You are now eligible for Head of Household filing status, for the Child Tax Credit, and for the Earned Income Tax Credit with children.

² We do envision potentially using such nearly-automated returns in specific corners of the tax system, for limited populations and use cases. More on this in [Section 14.3.3](#).

³ We are far from the first to argue this case. A 2003 Treasury report provocatively entitled [Return-Free Tax Systems: Simplification is a Prerequisite](#) explores similar dynamics and reaches similar conclusions.

⁴ The brief period when the Child and Dependent Care Tax Credit became refundable under the American Rescue Plan Act led to a Green Book recommendation to introduce information returns for child care payments. One of us provided technical assistance during the period when the Build Back Better Act contemplated extending the refundability provision, and, leveraging recent first-hand experience with child care providers and data systems, argued strenuously that the introduction of an information return requirement would be disruptive, costly, and politically disastrous.

The government does not know about this new child in your household, and would calculate your taxes incorrectly by potentially many thousands of dollars.

- Last tax year, you and your children moved out of your parents' home, where you had been living, and where the kids' grandparents paid the basic household expenses. You are now eligible to claim Head of Household filing status, drastically changing many aspects of your tax return, and potentially increasing your refund by over \$1,000.
- Last tax year, you began taking postsecondary courses at a local community college, but due to one of the exceptions, the school does not issue you Form 1098-T. You are now eligible for \$2,500 from the American Opportunity Tax Credit that the government does not know about.
- Last tax year, you started looking for work, and began paying your neighbor to watch your kids during the day. You are eligible for potentially thousands of dollars from the Child and Dependent Care Tax Credit.
- You have three children, and last tax year, you earned \$15,000 in W-2 income, but also \$300 in gig economy income not reported on any 1099 forms.⁵ Ignorant of this income, the government would underestimate your EITC by \$135.
- Last tax year, you moved from Nebraska across the border to Colorado in February. As a Colorado resident for 11 months, you are eligible for a 50% EITC match from the state for 11 months, potentially coming to over \$3,000. Neither the IRS nor state departments of revenue know when, or even that, you moved.
- You are an undocumented immigrant and held a W-2 job using an incorrect Social Security Number. The tax agencies do not have correct information on who you are in order to issue the appropriate refund.
- *(In New Jersey, or states with analogous programs)* Last tax year, you moved from campus housing (which is not subject to property tax) into a rented home where you pay \$20,000 in rent. You are newly eligible for a \$3,600 property tax deduction, which could save you hundreds of dollars in taxes.⁶
- You just made a contribution to an IRA, which allows you to take a deduction on your federal taxes. Because you just made the contribution, it has not yet been reported to the IRS on Form 5498. An automated tax return would not contain this deduction, causing you to overpay your taxes by perhaps hundreds of dollars.

These examples are limited only to cases where the missing data causes the taxpayer to receive too *small* a refund; there are also plenty of other cases where a taxpayer would underpay their tax.

And this is to say nothing of issues caused by late or erroneous 1099s or W-2s. Suppose the IRS automatically issues you an accurate refund — except it calculates the refund on April 1 without taking into account a 1099-INT reporting \$8,000 of interest, which ultimately was

⁵ \$300 is below the threshold at which payers are required to issue a 1099.

⁶ This example is calculated using New Jersey's Property Tax Credit/Deduction, but similar dynamics apply in a number of states.

reported to the agency late, on April 20. Now, a few weeks later, the IRS must come back and explain, unfortunately, you owe \$1,000 in taxes on that additional income.

In response to these challenges, advocates and policymakers sometimes raise a few rebuttals:

- *If we can't build return-free filing within the confines of the tax law, we should change the tax law.* Fundamentally rewriting the tax law might indeed make it possible to fully automate tax returns. But it should not be underestimated just how drastic such a rewrite would have to be. Taxes would likely have to be levied at the level of the individual rather than the family, to reduce the impact of unobservable family structure changes on deductions and rate calculations.⁷ The reforms would also probably require dismantling dozens or even hundreds of federal and state provisions most reform advocates support — and these sweeping changes would be required not only at the federal level but in all 42 income-tax states. To call this a daunting legislative agenda would be a wild understatement, and it could easily do more harm than good for the average American.
- *If there is data we do not currently have, we should improve reporting so that we do get it.* Tax agencies should certainly look at ways to expand reporting, so as to make life easier for taxpayers. But advocates should be clear-eyed about the monumental expansion of reporting required to fully automate tax returns. Considering only some of the examples above, full automation would require wholesale reporting of household rent paid, small-scale child care arrangements, family relationships and residency, immigration status, and much more, to federal and state authorities — and the removal of all exceptions under current requirements. This would not only be a serious technical challenge, requiring first-priority status within the federal and state tax agencies, but would massively change Americans' expectations and experience of privacy. Even the most pro-government enthusiasts might well have concerns about what a government with such unbridled access to Americans' information could do with that data.
- *Even if we can't automate everyone's returns, we should automate returns for those taxpayers whose returns we can automate.* Advocates may then concede that there will remain some data that some taxpayers must provide — but there will also be some taxpayers for whom no special cases apply, and the government could do away with tax filing *for those taxpayers*. In 2023, [Goodman et al](#) published a paper estimating that the IRS could fully accurately pre-populate a return for 40% of the population — why not, then, at least automate those 40% of returns? The trouble is: *who those taxpayers are* is the precise thing the government does not know. The government does not know which taxpayers have education expenses, family

⁷ U.K. taxes, for example, are levied on the individual rather than the family, allowing for much greater automation. But U.S. taxes are imposed at the level of a tax unit, whose definition is driven by marriages, residence, and relationships, all of which are not observable by the government. The tax unit definition drives the deductions allowed, the tax rates imposed, and often the calculation of various credits and other provisions.

structure changes, or moves across state lines that they need to report. Indeed, if the government knew these things, then we would not be here discussing what to do about taxpayers with unknown education expenses, family structure changes, or cross-state moves. In other words, the authors do not find that the IRS can fully automate returns for 40% of taxpayers; they find that, if the government fully automated *all* returns, they would be right 40% of the time. What the government *could* do is marshal the information it has, and then confirm the taxpayer does not have information to edit or add. In so doing, taxpayers would be advised they should *not* accept the government's best-guess data if it's wrong. And, if taxpayers *do* have additional information, the process would not kick them out of this "automated return" pathway, but rather allow them to *add that information*, taking advantage of the other automated data.

But this functionality — the government providing as much information as possible and requiring a taxpayer to go through and add/edit where needed — should sound familiar: it is precisely the endgame of Direct File we already described in the previous section.

There is no meaningful difference between asking a taxpayer to review their fully automated return on one hand and asking them to prepare a Direct File return with heavy pre-population on the other. **These are one and the same product.**

And this is the principal insight we hope policymakers and advocates of all stripes can take away from this discussion: **no matter whether you are a maximalist enthusiast of return-free filing, or a skeptic of any such reform, the path forward is to include as much automation in Direct File as possible, rather than building a parallel and more-automated filing pathway.**

- For advocates of simplified filing, who want to end filing as we know it: building a separate fully-automated filing product only for taxpayers without data to add/edit will just create a new headache when filers must determine which product to use, and will likely lead many to leave money on the table when they pick the wrong one. Advancing pre-population in the context of Direct File avoids these outcomes.
- For skeptics, who are concerned pre-population might engender tax noncompliance: advancing pre-population in the context of Direct File is the best way to ensure that pre-population is used in an environment where taxpayers can maximally understand what is being pre-populated and why, and understand their responsibilities to update incomplete or incorrect information.
- For technocrats concerned about the availability of information return data: advancing pre-population in the context of Direct File means that data availability can be treated iteratively. There is no need to wait until *all* of the data is available; we do not need to wait until we reach the ever-elusive world that a return can be automated wholesale. Direct File can simply be made available with the data that exists, and the user experience can improve over time as more data becomes available earlier in the year.

A few important points follow from this exercise:

- The outright eradication of tax returns is not possible without drastic changes to current tax law and the government ecosystem of the contemporary United States.
- To the degree it is possible to get closer to this goal of eradicating tax returns, the way to do it is by iteratively building on the automation and pre-population in Direct File. How far we can get toward the asymptotic goal of eradicating tax returns is not a question that has to be addressed up front; it is a boundary that will continue to advance as Direct File grows, and the tax law and reporting ecosystem evolves.
- We believe it is vital that pre-population remains a fundamentally *opt-in* experience: when they use Direct File, taxpayers will have to actively elect to use pre-population. Because any pre-population will be incomplete and require taxpayers to confirm, correct, and augment their data, it is critical that taxpayers understand what they are choosing to do, and do it on their own terms. Otherwise, it will be hard to maintain the critical principle that taxpayers are, fundamentally, responsible for the accuracy of their own returns.

2.1.3 Pre-population, pre-fill, data import, simplified filing, and return-free filing — defining our terms

To put it mildly, pre-population has been the subject of significant controversy. Tax administrators are concerned that returns pre-populated with incomplete information could encourage taxpayers to file incomplete returns, reducing tax compliance. Some taxpayer rights advocates are concerned that taxpayers would be hesitant to correct their pre-populated return, thereby causing them to leave money on the table. Perhaps most important of all, private tax preparation companies have opposed the implementation of functionality that they see as giving the IRS a leg up on their products. Amid all of this, various actors have tried to mollify the various factions, assuring them that this or that proposal will not trigger their broader concerns. All this controversy and jockeying has led to a proliferation of different terms, all generally connoting “using IRS data to streamline tax filing,” but with slightly different implications and baggage:

- *Automation* usually describes the use of IRS data to require less of the taxpayer and thereby streamline the filing process — whether by showing the taxpayer information to review, by retrieving answers behind the scenes and not asking questions at all, or by correcting taxpayer error after filing. (More rarely, automation has been used to mean the eradication of tax returns entirely.)
- *Pre-population*, likewise, usually describes the use of IRS data to require less from the taxpayer. Narrowly speaking, the term itself implies a more limited scope of change in the taxpayer experience: the data is used to create draft answers, but not used to hide questions wholesale or draw inferences without taxpayer involvement. By this distinction, Direct File prior-year AGI retrieval in 2024 would be mere pre-population, whereas One-Step Signature (which, based on logged-in status,

waived AGI retrieval altogether⁸) would be automation. In practice, though, most writers or policymakers see *pre-population* as encompassing both.

- *Pre-fill* is effectively a synonym for pre-population (and, by proxy, automation). In some circles, pre-fill was thought to be a more concerning term than pre-population, suggesting that taxpayers would barely even review the pre-filled information, whereas pre-populated information would be offered more as a suggestion. In practice, there has become very little distinction between these two terms.
- *Data import* was the term the Direct File team used to describe its pre-population functionality in 2025: Direct File allowed you to import your data from the IRS. The team used this word intentionally, intending to avoid hornets' nests around terms like pre-population or pre-fill. In practice, this linguistic distinction did not really accomplish much.
- *Pro forma return* tends to indicate a situation where the tax agency completes a tax return for the taxpayer, and while nominally a taxpayer must approve it, making adjustments is very rare. In the vast majority of cases, taxpayers will just say “yes” to what has been prepared. This is along the lines of the system in Finland or Estonia.
- *Return-free filing* is used either as a synonym for *pro forma return* or perhaps even one step farther — a world in which there is no longer anything in the tax system that resembles a tax return, because precision withholding has made the process of an annual tax reconciliation entirely unnecessary. The amount you paid throughout the year is exactly what you owe, and there is nothing to reconcile. This is along the lines of the system in the U.K.

We contend that these minute distinctions, at this point in the story of tax filing reform, cause more confusion and harm than good. Fundamentally, ***pre-population, pre-fill, automation, and data import should all be understood as synonyms, for the process of iteratively improving on Direct File so that it requires less and less of taxpayers. Return-free filing and pro forma returns are extreme versions of this process that are not per se achievable in the United States, but are the north stars that the process of iteration asymptotically approaches.*** In this report, we usually refer to ‘pre-population’ or ‘automation,’ which together share the goal of ‘ending filing as we know it.’

2.2 Data sources and limitations

While all pre-population entails using IRS data to streamline the filing process, IRS data is not a monolith. Different pieces of data are available to the IRS from different sources, in different ways, at different times. Sections 2.2.1-6 explore the main data sources, and some of the important issues inherent to each of them. Section 2.2.7 discusses a technical issue affecting all data sources.

Often, the discourse can become fixated on just the first of these data sources — information returns like W-2s and 1099s. But, as we will see, these are only a part of the

⁸ Behind the scenes, these returns were treated for IRS processing purposes as if they were filed by a tax preparer with an e-filing authorization on file.

picture, and other types of data can be just as valuable in streamlining tax filing. Above all, we hope this discussion elucidates the vast and multi-faceted project that is data pre-population, and how it can advance along multiple different dimensions, with different types of benefits for different populations of taxpayers.

2.2.1 Third-party data, or information returns (W-2s, 1099s, 1095s, 5498s)

Income (plus withholding and other relevant tax details) is reported to the IRS by employers/payers on Forms W-2 and 1099. These forms — plus 1095s, 1098s, and 5498s, relating to health care coverage, education expenses, and retirement contributions, respectively — are called information returns. The bulk of the discourse around pre-population concerns this data.

Direct File built pre-population in 2025 for W-2s, 1099-INTs, and (partially) 1095-As in 2025, meaning that the legal and basic technical aspects of information return pre-population are essentially a solved problem, though there is much scaling to do.⁹

Most of these forms contain predominantly federal information, but they also often contain additional fields with state-specific information, related to the amount earned in each state (and, if applicable, locality), withholding paid thereof, and sometimes other state-specific details.¹⁰ While this data is not literally used in the preparation of a federal return, it is usually transcribed into a taxpayer's return during the federal portion of the filing experience, as it will be needed when the taxpayer gets to their state return. This way, taxpayers deal with their income documents just once, not twice, during the filing process.

2.2.1.1 W-2

Forms W-2, reporting wage income, are by far the most important of these information returns, with about 76% of all returns reporting at least one W-2.¹¹ Direct File already built pre-population of W-2s for filing season 2025.

Unlike other information returns, Forms W-2 are not reported directly to the IRS. Instead, the data is sent to the Social Security Administration, which forwards it to the IRS. But **SSA normalizes the data en route**: (1) it aggregates some of the data in Box 12, regarding contributions to retirement plans, which need to be provided separately to prepare a proper return, and (2) it drops data from Boxes 14-20, which are usually used in the calculation of

⁹ The idea of third-party pre-population also has precedent in the private tax preparation ecosystem, where tax prep companies have built integrations with payroll processors and financial institutions which allow taxpayers to import and pre-populate some of their W-2, 1099-INT, and 1099-R data, among other forms. That these integrations are with private systems rather than the IRS means their coverage is less complete. But the basic precedent is well established in the tax ecosystem.

¹⁰ For example, Form W-2 may contain documentation of contributions to pension plans that are deductible under state, but not federal, law.

¹¹ Based on the number of Forms 1040 reporting an amount in Line 1a, according to [IRS Publication 4801](#).

state and local taxes. **This means the IRS’s W-2 data is missing critical information.** In most cases,¹² taxpayers who used the W-2 pre-population functionality in 2025 needed to retrieve their own copy of their W-2, anyway, to add this missing information — which badly limited the functionality’s usefulness. Fixing this gap is critical to Direct File’s future.

We believe this issue can be solved administratively: with good intra-administration collaboration, SSA can and would share the required data. While SSA faces some technical challenges due to its fixed-width data systems, these issues are surmountable, and there are no statutory barriers to data sharing. Indeed, most of this data is currently removed only because the IRS historically had no need for state-specific data in Boxes 14-20; IRS simply must now articulate the use case for this data. **We believe working with SSA to remedy this data pipeline is a far faster route than alternative paths, like pursuing a legislative fix, bypassing SSA and having W-2s reported straight to the IRS, or working with states to independently pre-populate state-specific boxes from their W-2 data (see Section 2.2.6).**

The routing through SSA also slightly delays the IRS’s receipt of W-2 data. W-2s are due to SSA on January 31, but it can take a couple of weeks for that data to be available at the IRS. This process, too, could likely be sped up administratively as systems are modernized and with the articulation of a clear use case.

2.2.1.2 Other information returns

Then there are the various other flavors of information returns. These include:

- 1099s reporting income from non-wage sources (e.g., interest (1099-INT), dividends (1099-DIV), unemployment benefits and some other government payments (1099-G), gig economy income (1099-NEC));
- Education-related expenses that may have tax implications (e.g., qualifying tuition fees (1098-T));
- Contributions to tax-privileged accounts (e.g., contributions to ABLE accounts (5498-QA), contributions to IRAs (5498));
- Distributions from tax-advantaged accounts which may have tax consequences (e.g., retirement accounts distributions (1099-R), 529 distributions (1099-Q));
- Health care coverage (e.g., marketplace coverage (1095-A), employer coverage (1095-C)).

Eventually, getting information from all of these forms — of which there are dozens — is critical to Direct File. It’s also a tall order, and could take many years.

In 2025, Direct File offered pre-population of 1099-INT, and retrieved data on *whether* the taxpayer had a 1095-A, though not the information on it. The mere fact of the form’s existence allowed Direct File to warn a taxpayer that they must go through the Premium Tax Credit functionality.¹³ And, technically, it is far easier to detect a form’s existence than to

¹² Taxpayers in states without income tax and without relevant data in Box 12, would not have needed to transcribe their own data.

¹³ Otherwise, their return would be rejected by the IRS e-file system’s automated business rules.

make full use of all the data on it. This dynamic could be valuable as Direct File iteratively imports data from more and more forms — the easier-to-incorporate presence of forms can be a valuable first step. In the case of other information returns, the presence of a form can help Direct File determine if taxpayers have sources of income that are disqualifying for a still-limited Direct File scope.

2.2.1.3 Timing of information returns

As summarized in the [2025 Direct File report](#), information returns have a range of deadlines, some of which are so late that they pose a serious problem for the usefulness of pre-population. Other than W-2s and 1099-NECs (gig economy income), which are due January 31, most 1098s and 1099s are only due to the IRS by March 31 (or February 28, if e-filed). **Moving the due dates to January 31 for these forms is essentially a prerequisite for a more aggressive pre-population regime.** Unfortunately, this will require a statutory change to 26 USC 6071(b), which establishes the March 31 deadline.

This is to say nothing of W-2s and 1099s that arrive after the deadline, which are also a serious issue. As the [2025 Direct File report](#) lays out, 1099-INTs are due to the IRS by March 31, but the average date of receipt is April 13, and 20% of forms are still outstanding by May 22. **More rigorous enforcement of these information return filing deadlines is also a prerequisite for a more aggressive pre-population regime. This may have to include improvements in the systems that payers use to file these forms.** There is some precedent for the idea that a more aggressive enforcement regime alone can actually improve the timeliness of information return data: according to non-public presentations at tax administration events, some states have seen significant impacts from implementing concerted programs to enforce information return deadlines.

Of course, accelerating deadlines and increasing enforcement could raise objections from employers and other institutions subject to those requirements, and their political allies. We believe the project is worth the risk, though, not least because **the benefits do not end with pre-population, and are not restricted to one side of the political spectrum.** Faster reporting (indeed *far* faster than that envisioned here) will be needed to effectively enforce SNAP and Medicaid work requirements, passed in H.R. 1 in July 2025 (see [Chapter 11: Real-time taxes](#)). **Faster reporting will also empower the IRS to more rigorously detect and crack down on errors and fraud in individual income tax returns**, a longtime conservative goal in Congress, currently badly stymied by the fact that the IRS does not have the data it would need to dispute many tax filers' self-reported income. Moreover, payers already must furnish 1099s to *taxpayers* on January 31. How much additional burden is it really to align the deadline for sending to the IRS with the deadline for sending to taxpayers?

There is also the possibility that data forced to arrive earlier will simply be lower-quality data, which must subsequently be corrected — which would do more harm than good. **More research could be helpful to confirm that earlier reporting is feasible for all information return types, and what kind of allowances are needed to ensure that proper enforcement**

for the majority of returns does not cause adverse consequences for a more complex minority.

Meanwhile, there is the issue of 5498s, reporting contributions to tax-privileged accounts like 529s, ABLE accounts, and IRAs, which can have tax consequences. These forms are due to the IRS by May 31. But the deadline cannot be simply moved up; taxpayers do not actually have to make the contributions to these accounts until April 15. So a 5498 sent to the IRS by January 31 would of course be missing a contribution made April 1, which should be reflected on a return filed the same day, or a week later. It is not obvious how to resolve this issue, and it will take additional user research and design.

2.2.1.4 Errors in information returns

The above sections note that Direct File taxpayers need the option — and indeed the responsibility — to correct data pre-populated from their information returns if it turns out to be wrong. Making these corrections, though, could easily create problems. If a taxpayer significantly edits their income amount, for example, it is quite possible the return will get flagged on account of that discrepancy and subject to enforcement actions. If the taxpayer edits their withholding amount, on the other hand, the discrepancy will almost certainly be immediately detected and “corrected” back to the previously reported value, taxpayer edit notwithstanding.

As such, in a robust system, edits taxpayers make to pre-populated data will likely need to be connected back to a process of identifying and correcting the information return source data. Depending on the nature of the correction and the error, this will entail the involvement of the payer who issued the information return, plus the notification of any state and local tax agencies to which the same information return was reported. The importance of triggering this process would depend on the nature and magnitude of the edit. More types of edits and smaller-magnitude edits would become more relevant over time, as modernized IRS systems implement lower tolerances for discrepancies.

The IRS would also need to design a process for the case where a return was filed on the basis of a pre-populated information return that was later amended via a superseding corrected form. For example, suppose a W-2 is filed on January 20, and the taxpayer pre-populates this data and files (without edits) on February 20; but the payer then files an amended W-2 on March 20 significantly editing the taxpayer’s income and withholding. There should be some automated process to alert the taxpayer and expedite a return amendment.

2.2.2 Data from taxpayer account or identity proofing

W-2s and 1099s relate to income received by the taxpayer specifically during the tax year of record. Meanwhile, the IRS has much information about a taxpayer on an ongoing basis, often tied to a taxpayer’s account. **Some of this data is much more esoteric — which, by the same token, makes it especially valuable, since it can be challenging for taxpayers to track down.**

Some of these are items Direct File implemented in 2024 or 2025. (In some cases below, Direct File automated data items whose provision is usually used as verification of a taxpayer's identity. This worked because Direct File taxpayers had already cleared IAL2 identity verification via ID.me, a higher bar (see [Chapter 5](#)), which made this additional verification unnecessary.)

- *Prior-year AGI or self-select PIN.* To verify their identity on e-filed returns, all taxpayers must provide their prior-year income or the PIN they selected when filing their last return. Many taxpayers get these items wrong, causing their returns to be rejected. The IRS, naturally, has easy access to this information, and the extra validation step is unnecessary for identity-verified taxpayers. Direct File allowed taxpayers to import their prior-year AGI in 2024. In 2025, Direct File got rid of the unnecessary middle step, calling the functionality “One-Step Signature.”
- *Identity Protection PIN.* IP PINs are six-digit numbers issued annually to taxpayers who have previously been the victim of identity theft, or who voluntarily enrolled in the program. They must be provided on the return. In practice, some taxpayers have lost track of their IP PINs, or, especially in the case of intermittent filers, [may be unaware they are enrolled in the program at all](#). Again, the IRS has this information, and the extra validation is unnecessary for verified taxpayers. Direct File pre-populated this data in 2025.
- *Biographical information* — that is, full name, address, email address, birthday, SSN, etc. This information is either stored in the taxpayer account or comes directly from the identity proofing service. Direct File pre-populated this information in 2025.

Other items are rarer and more specific to smaller populations, but could be just as valuable for them. Direct File did not implement these items, but they would be natural extensions of the functionality already built.

- *Estimated tax and extension payments.* The IRS knows exactly what quarterly estimated payments or extension payments taxpayers made over the previous year, and would correct this information post-filing if the taxpayer entered it wrong, anyway. That the IRS knows the status of these payments is so obvious to taxpayers that it would be a meaningful public relations benefit to be able to automate this part of the process. Interestingly, there was also some evidence from the early Direct File experience that asking about estimated payments is not just a nuisance for taxpayers who made them, but a source of meaningful confusion to taxpayers who *didn't*, who often do not know what estimated payments are, and sometimes erroneously enter other payments (withholding or property taxes, for example) in this field.
- *EITC/CTC bans under 32(k) and 24(g).* Taxpayers who are found to have improperly claimed the EITC or CTC are banned from receiving it in the following two or ten tax years (depending on the severity of the improper claim), under sections 26 USC 32(k) and 24(g), respectively. Moreover, even after the ban is over, taxpayers must file Form 8862 the first year they start claiming the credit again. Currently, Direct

File must ask highly esoteric questions about whether the taxpayer is subject to a ban, and if they have already filed the 8862. The IRS already has this data.

- *Taxpayer communication preferences.* Taxpayers can use their tax returns to opt into receiving communications from the IRS [in other languages](#) or [alternative media formats](#). Currently, Direct File asks if taxpayers want to set such a preference. Instead, Direct File could display the current setting and allow taxpayers to amend it.

2.2.3 Data from other current-year returns

One category of data not often discussed in the context of pre-population is data from other taxpayers' current-year returns, in the IRS Master File. This data figures into the filing process today through the use of MeF reject codes after a return is filed, but a better taxpayer experience would incorporate it into the filing process (provided these changes can be implemented without opening pathways for fraud or unauthorized disclosure). For example:

- *Taxpayer claimed as dependent.* Taxpayers must report if they are claimed as a dependent on another return, which has far-reaching tax implications. Currently, if a taxpayer has been claimed but does not say so on their return, the return is rejected upon submission. Direct File could instead warn taxpayers who have already been claimed on another return well before submission.¹⁴ (This piece of data cannot be used to automate dependency entirely, since the return claiming this taxpayer may not yet have been filed.) The dynamic is similar for the closely related question of whether a taxpayer has been claimed as an EITC qualifying child.
- *Taxpayer already filed.* Currently, if a taxpayer has already filed a return, a second e-filed return would be rejected. A better taxpayer experience would warn the taxpayer up front, before they prepare a whole return.
- *Dependent already claimed.* Currently, if a taxpayer claims a dependent who has been claimed on another return, the return is rejected. A better taxpayer experience would warn the taxpayer upon their entering the dependent on the return. (Unlike in current practice, taxpayers should also be able to e-file conflicting dependent claims, a topic discussed further in [Chapter 12: claiming dependents](#). In a world where taxpayers are warned about the conflict before filing, the return itself could be used to explicitly kick off the resolution process.) This item is potentially more challenging from a fraud and privacy perspective than the previous two.

Unlike with information return and account data, Direct File had not yet examined the technical or legal implications of automating this data, a topic that would have to go through IRS counsel. **There is a strong precedent for sharing this data: taxpayers already get this kind of feedback in the form of MeF reject codes.** Using such data during the Direct File return preparation experience basically just moves the disclosure of such data

¹⁴ Of course, for privacy reasons, Direct File could not reveal *who* has already claimed the taxpayer, simply that someone has.

forward in time — although care must be taken to avoid creating new avenues for fraud or unauthorized disclosure.

2.2.4 Data from prior-year returns

Even taking full advantage of information returns, tax account data, and data from other current-year returns, there is still a lot of information missing to complete a tax return — information not knowable from these other sources. This includes critical information about a taxpayer's family structure, and personal details that the government is not in any reasonable position to observe for the current tax year. All together, this data adds up to a large portion of the screens a taxpayer sees in Direct File, especially in the Family and Household section.

But, these are all data items that are quite likely to remain constant year over year. Using prior-year returns to provide a starting place for taxpayers could save them significant time on their current-year returns. Seeding a return with prior-year data is standard practice in private tax software, and taxpayers indeed cite this prior-year rollover as a critical reason they remain wedded to their existing tax prep solution. Of course, an IRS solution has the key advantage that prior-year data could be carried forward regardless of how a taxpayer filed the previous year.¹⁵

Important uses of this data include:

- *Marital status (and, if applicable, spouse information).* Direct File could pre-populate information about the prior year marital (and filing) status, asking taxpayers just to confirm key facts, rather than start from scratch. In the case that a spouse must be added to the return, this would obviate the need for a lot of tedious data entry — often about 5-10 screens in Direct File.
- *Dependents/household information.* Like marital status, pre-populating information about household members would save a lot of data entry — usually at least 10-12 Direct File screens per household member. Again, Direct File would need to ask taxpayers to confirm a few key facts still hold, so that pre-populating data does not become an invitation to disregard life changes. In the case of dependents in particular, Direct File would also naturally take into account the passage of time, so that, for example, a (student) child aged 23 the prior year would not be claimed as a qualifying child this year.

¹⁵ Direct File would be able to carry forward more prior-year data from former Direct File users than from other tax filers, since not every data item a taxpayer provides is explicitly encoded in the tax return; data might, instead, be used to make a calculation or determination, with only the output of the calculation/determination included on the return. (For example, a single parent with a child may not be eligible for Head of Household filing status because they do not pay half the household expenses for the child's home. But that fact itself is not included on the tax return; all that is included on the tax return is the filing status Single, which could have been selected for any number of other reasons.) Direct File would, then, have to navigate a two-tiered situation of prior-year pre-population, where some pre-population is based on prior-year Direct File returns, and is more robust than the pre-population based on another return.

- *Disability status.* Taxpayers must answer a variety of questions about disability and blindness status, for a variety of tax provisions, which can be pre-populated with prior answers.
- *Immigration status.* Taxpayers must answer a variety of questions about the immigration status of themselves and their families, for a variety of tax provisions. In principle, this is data that could be retrieved from other federal data sources, but privacy considerations will tend to militate strongly in favor of not using this data, and instead pre-populating with prior answers.
- *Miscellaneous items.* There are countless additional small items on the tax return where prior-year data is a reasonable guess for current-year information; third-party designee and bank information are two examples.

The above categories are all cases where there can be a reasonably strong presumption that data remains the same year over year. Then there are other **cases where prior-year data might be used to prompt a taxpayer to take a closer look at a particular aspect of the return.**

- *Credit claims.* If a taxpayer in a previous year claimed a provision like the Child and Dependent Care Credit, the American Opportunity Credit, or the Educator Expenses Subtraction, the IRS will not usually have the full information to automate the credit claim in the current year, but Direct File can nudge the taxpayer: “Last year you claimed \$500 in educator expenses. Did you have educator expenses to claim again this year?”
- *Income sources.* If a taxpayer in a previous year reported a certain category of income, it might be useful to prompt the taxpayer to check carefully if they had this category of income again. This might be especially useful in a world where income pre-population is more comprehensive, but there are persistent issues with late-arriving information returns. Suppose for example a taxpayer has historically usually reported interest income, but there are no 1099-INTs yet on file for that taxpayer. Direct File might say: “Usually, you report \$200-300 in interest income on your taxes. Did you have such income again this year? We don’t see any forms reporting this income, but it’s possible they haven’t arrived yet.”

Prior-year data might also be useful to improve transparency for taxpayers seeking to understand their taxes. In practice, taxpayers often expect their refunds to be about the same size year over year, and wonder what went wrong if the number is significantly different. Direct File already has a robust page explaining how a taxpayer’s refund or liability was calculated. It might be possible to incorporate prior-year data into this functionality, and **explain the principal changes between last year and the current year’s return.**

Finally, all of the above uses involve the use of the taxpayer’s *own* prior-year returns. **But there is also the possibility of taking into account information from other prior-year returns.** For example, suppose a taxpayer begins to claim a dependent who has previously been claimed by another taxpayer for many tax years; or a taxpayer reports not being a

dependent, even though they have been claimed as a dependent in many previous tax years. In this case, Direct File might use these patterns to warn the taxpayer about a potential error, though of course would not block the taxpayer from making the claim.

Like using current-year data from other returns, Direct File had not yet examined the technical or legal implications of automating prior-year data, a topic that would have to go through IRS counsel.

2.2.5 Data from SSA

The IRS receives birth records (with parent information) from the Social Security Administration's Enumeration at Birth program, which can be used to suggest new children that a taxpayer may want to claim. Direct File might say for example: "According to data from the Social Security Administration, you had a child this year. Would you like to add [child name] to your return?" The IRS also accesses death records from the SSA Death Master File, which can abbreviate a variety of — potentially emotionally difficult — screens regarding the death of a loved one.

2.2.6 Data held by the state tax department

The previous sections discuss IRS data (or data acquired centrally, by the IRS). But, as discussed further in [Chapter 4: Direct File and state taxes](#), the U.S. has a federalist tax system, with state (and local) returns often looming just as large as the federal return. Generally speaking, taxpayers are required to prepare their federal return first,¹⁶ and state returns are based on the data collected and/or transcribed while preparing the federal return.¹⁷ So to the degree a state return is "pre-populated," it is with the income, family structure, and credit/deduction data from the federal return — whether that data was initially entered by the taxpayer or is itself pre-populated.

But what about data the state tax department itself has? Could this data be valuable in the pre-population of the broader federal-state Direct File experience?

Most state tax department data is duplicative of information that will already be documented during the completion of the federal return, and should *not* be independently pre-populated from state databases into a state return. For example:

- *Information returns.* W-2s and 1099s are reported to state tax departments, just as to the federal government. But by the time a taxpayer reaches their state return, they will have already entered or retrieved their information return data while completing their federal return. Not only would re-pre-populating W-2 data on the state return be unnecessary, it could actually be damaging, as it might induce a taxpayer to

¹⁶ Most state returns explicitly request data items from a federal return. In the case of taxpayers who do not file a federal return, many states require taxpayers to calculate figures exactly as if they had completed that federal return.

¹⁷ Puerto Rico is a notable exception to this rule, where the federal return requires information about the Hacienda tax.

inadvertently introduce a discrepancy. Suppose the W-2 issued by an employer contained an error, which the taxpayer caught and resolved on their federal return. It is this corrected version that the taxpayer ought to use on their state return; but pre-populating the state copy in the state functionality could induce the taxpayer to import the wrong version, which is incorrect and will not match their federal return or the income information imported from it. The already-completed federal return, in other words, is probably a more reliable source of information returns than the state department of revenue.

- One nuance here has to do with the gaps in information return data discussed in Section 2.2.1. Some states may receive some information returns before the IRS, and probably all states store the state-specific W-2 data that SSA currently strips out (Section 2.2.1.1). Some have argued that pre-populating information return data from state sources would elegantly solve for gaps in federal data. We contend, though, that this would be a wild goose chase. Even if a few states have better or faster W-2 data and could easily plug this data into their state return functionality (itself a big *if*), most states likely do not, and solving the data issues in these states would be harder than solving it federally. In any case, it would entail solving a problem 42 separate times that could instead be simply solved once. For whatever the challenges to the SSA solution discussed in Section 2.2.1.1, making up the difference with state data will almost certainly be harder.
- *Family structure data from prior-year returns.* States have information about dependents and family structure from prior-year returns, just like the IRS. But, again, by the time the taxpayer gets to the state return, they will have already had to enter and confirm their up-to-date family information in their federal return, and re-pre-populating would only risk creating new discrepancies.

There are a few examples of data items that the state is in a position to know and the federal government is not. For these items, it might in principle be valuable to build pre-population flows using state data to streamline the state filing process. Consider, though, that these are much lower-level items than the much larger range of items discussed in above Sections 2.2.1-2.2.5, and it is unclear whether the juice will be worth the squeeze to build these flows, which would involve substantial technology work, plus possibly even a whole separate identity verification flow to validate the state data requests.¹⁸

¹⁸ States would need to validate that taxpayers are who they say they are before releasing such sensitive tax data to the user. If states and the IRS used a common identity provider (such as Login.gov, or ID.me), then perhaps this validation would be trivial enough for the end user, though still could require some technical complexity. But if, as at present, states departments and the IRS do not share a common identity solution, end users would have to separately validate to a state identity system — an onerous process, akin to requiring taxpayers to go through ID.me verification twice. Going through it once, as discussed in [Chapter 5: Direct File and identity verification](#), was often trouble enough.

- *State estimated tax and extension payments.* Taxpayers make estimated and extension payments to the state just as they do to the IRS, and this data could be retrieved from state sources.
- *Prior-year information for state provisions.* Just as Direct File might prompt taxpayers with prior-year information to take a look at certain tax provisions or suggest information added last year, states could do the same for state-specific credits, additions, subtractions, and deductions, eligibility for which is not reflected on a state return, e.g., Renters Credits.
- *Explanation of refund compared to prior year.* Just as Direct file might use prior-year information to explain this year's refund and put it in context, states could do the same.
- *State credit bans.* If any states have credit bans along the lines of 26 USC 32(k), these could be pre-populated into state returns. That said, we are not in fact aware of states actually having these provisions on the books.
- *Property tax data.* Many states have income tax credits or deductions to defray the cost of property taxes. These provisions are onerous for taxpayers, and pre-populating information about property taxes paid would critically help streamline the significant complexity of claiming these credits. That said, in at least most states, property taxes are managed independently of the state income tax department, and there may not be any easy way to pre-populate this (usually local) property tax data. As such, state tax pre-population may or may not be helpful in these cases.

In short, there are some possible uses for pre-populating state income tax data, but they are relatively few and far between. As a result of these dynamics, while pre-population of state tax data might be a nice-to-have, it is certainly not the game-changer that federal pre-population is. Tax administrators must determine whether the technical effort to build integrations and the taxpayer experience implications of a second verification flow are worth it.

2.2.7 Technical issues with non-primary taxpayers

All of the pre-population discussed above entails revealing to the Direct File user highly sensitive taxpayer data. Naturally, taxpayers using this functionality need to clear a high bar of identity verification, along the lines of the standards already being enforced in Direct File (see [Chapter 5: Direct File and identity verification](#)). It would be inappropriate, not to mention potentially illegal, for the IRS to reveal income and prior-year return information to taxpayers who have not confirmed they are who they say they are.

A complicated wrinkle here, though, concerns data of spouses. In the current Direct File architecture, only the primary taxpayer clears identity verification. The IRS cannot display to this logged-in taxpayer the data that belongs to anyone whom they might say is their

spouse, unless that (alleged) spouse consents.¹⁹ Since both spouses have their own information returns, spouses may have had different prior-year returns, and most questions on a joint return concern both spouses, solving this problem is critical to the success of data pre-population.

The clearest solution would be for both adults to have separate IRS accounts, and the spouse would log into their own account, consenting there for their data to be exposed to the primary taxpayer. There might also be streamlined methods to authorize current-year data disclosure for married couples who have jointly verified in the past.

The spouse data problem is solvable, but will require careful user research and design work by an empowered team to come up with an elegant solution — and, of course, this solution will have to fit in with whatever identity verification policies apply to Direct File as a whole (see [Chapter 5: Direct File and identity verification](#)).

2.3 Limitations of Direct File functionality built to date, and expansion considerations

The existing Direct File pre-population functionality is summarized at length in [the 2025 Direct File report](#), where it is referred to as Data Import. As noted there and above, by its second year, Direct File was importing W-2s (with limitations), 1099-INTs, and the presence of 1095-As; biographic information from ID.me; and prior-year AGI and Identity Protection PINs from the taxpayer account. These were only all for the primary taxpayer, not for the taxpayer’s spouse.

An additional dynamic only partially explored in the 2025 report concerns the way that this pre-population functionality was built into Direct File. Direct File was built first for the experience where a taxpayer answers every question manually. Due to time crunches in delivering this functionality, pre-population was largely bolted on to the existing product structure, both in its technology implementation and in terms of the taxpayer experience. This is not to say that the taxpayer experience did not look different for the pre-populated path, or that the functionality was introduced haphazardly; the changes and the philosophy behind them are discussed at length in Appendix K of the 2025 report. But more would be needed to actually build a taxpayer experience that is “pre-population-first.”

¹⁹ This might appear tedious and counterintuitive, but consider the implications. If no proof were required, a bad actor could verify themselves to Direct File, report they are married to any arbitrary person in the country, and thereby view that person’s sensitive tax data. And even if the taxpayers are in fact married, it does not follow that the IRS should reveal to one spouse the other’s data, even if this might be quite normal in many marriages.

2.4 Next steps and recommendations

2.4.1 Expanding pre-population to include more data items

Expanding pre-population is largely a matter of simply empowering a Direct File team to continue the work that was happening in 2025. As pre-population expands, there are a few principles to keep in mind:

- *Pre-population is iterative.* As discussed in Section 2.1, a pre-population team’s job is not to jump to some new notion of an automated return, it is to deliver more and more automation each year, always prioritizing delivering smaller pieces sooner than waiting to deliver big chunks all at once. “Return-free filing” is simply the ending point of this process of pre-population iteration.
- *Pre-population is about more than just information returns.* The full roster of possible types of data to pre-populate are discussed at length above in Section 2.2, and a team working on pre-population should think broadly about what it can mean. The team should consider starting to work with IRS counsel early on about additional categories of data for pre-population (e.g., prior-year return data), as a new category may come with more legal review than simply a new form or new item within an existing category.
- *Prioritization.* As in the case of prioritizing new tax scope (see [Chapter 3: Direct File and tax scope](#)), new elements of pre-population should be prioritized based on: (1) number of taxpayers impacted, (2) complexity averted for those taxpayers (with a focus on items that save taxpayers time, items that improve accuracy, and items that might otherwise be showstopping barriers for new/intermittent filers), and (3) effort to implement. There is no magic formula here; a pre-population team should go screen by screen through the product, considering what can be waived, and how valuable it would be.
- *Detecting the existence of a form is easier than making full use of it* (as discussed in Section 2.2.1.2), and the pre-population team should consider applying this dynamic to its advantage in constructing a road map. Detecting the presence of a wide range of forms early on would be incredibly valuable in improving Direct File’s eligibility check functionality, which is currently hard to navigate.
- *Cross-team prioritization.* As discussed in [Section 3.5.3](#), capacity is to a degree fungible across different parts of the Direct File team. Direct File writ large can to an extent trade off expansion of pre-population against expansion of tax scope, or state functionality. We believe pre-population ought to be a high priority, but the details of those trade-offs will depend on circumstances and timelines.
- *Spouse data.* Creating an identity verification solution to pre-populate spouse data (Section 2.2.7) is a high priority.

2.4.2 Pre-population, Direct File architecture, and the taxpayer experience

A new Direct File team may have to do some foundational work reimagining how pre-population is built into Direct File. Although pre-population was always a goal of the product, tight timelines in the early years meant that the actual implementation of pre-population was somewhat jury-rigged, both in terms of the taxpayer experience and the underlying architecture (see Section 2.3).

In general, as pre-population was implemented in 2024 and 2025, the default was to show all screens — with then some fields pre-populated, or at best a handful of screens skipped, thanks to the imported data. As pre-population grows, though, it will be able to transform the taxpayer experience in more profound ways, something a team may have to consider from the ground up. Questions like this will become especially relevant as Direct File begins including pre-population from prior-year data, which cannot be included without *some* taxpayer review, but should feature a notion of taxpayer review that does not require going through all the pages a from-scratch taxpayer would. It would also be relevant to the income section, in which, rather than asking taxpayers to go through potential sources of income category by category, Direct File might display all known income and ask taxpayers to confirm it looks right.

Part of this reimagination and optimization entails the ordering of different data items. For example, currently, Direct File collects family and household information before income, which is important so that Direct File knows which individuals income must be collected from. But, one might want to do a basic review of income sources first, so that taxpayers ineligible for Direct File can be identified up front without spending energy answering a long and complex string of family questions.

2.4.3 Pre-population and state data

As discussed in 2.2.6, state tax departments have their own databases that could be used to pre-populate state returns — but, for the most part, these databases contain few data items that are not already present on the federal return. The Direct File team should work with state partners to determine if the gains from state pre-population are worth the technical effort and taxpayer experience implications. We suspect they likely are not, but the decision should ultimately be left to the implementation teams. If state pre-population is pursued (and, arguably, even if it isn't), federal and state teams should look for ways to reuse federal identity verification at the state level.

2.4.4 Information return availability and reliability

Improving the availability and reliability of information returns is central to the project of pre-population.

- *Collect W-2 Boxes 12-20 in full from SSA.* SSA and IRS must work together to ensure IRS receives the complete W-2, including (disaggregated) Boxes 12-20 (see Section 2.2.1.1). This is the most expedient and universal solution to the problem of incomplete W-2 data. SSA and IRS may also look at speeding up the time between W-2 arrival at SSA (January 31) and their availability at the IRS a couple weeks later.
- *Move up information return deadlines and improve enforcement.* To make pre-population viable, most information returns (including, e.g., 1099-INT, 1099-R, and 1099-G) need to be delivered to the IRS by January 31 (in line with the deadline for delivering those forms to taxpayers, and in line with the deadline for reporting W-2s and 1099-NECs to the IRS), and the IRS needs to more aggressively enforce the existing deadlines. Research is needed to ensure this enforcement does not lead payers to provide lower quality data. This would require statutory changes to 26 USC 6071.
- *Design for the case that not all information returns are available.* Even in the case of quick and decisive action, it would be some time before all information returns are available by early in the filing season. Direct File must be designed for the world in which taxpayers use Direct File before their information returns are fully available, with appropriate warnings about the need to add any missing information.
- *Create processes to flag and resolve likely errors in information returns.* If taxpayers edit their pre-populated information return data while using Direct File, this edit should kick off a process to review and potentially amend the information return data. The results of this process, which may include the involvement of the return issuer, should ultimately be reported out to relevant state and local tax departments as well.
- *Create an expedited process to amend returns when later information return data contradicts earlier data.* If taxpayers prepare their returns based on pre-populated information return data that is later altered by an amended information return, Direct File and the IRS should have an expedited process for updating the taxpayer's return and refund.