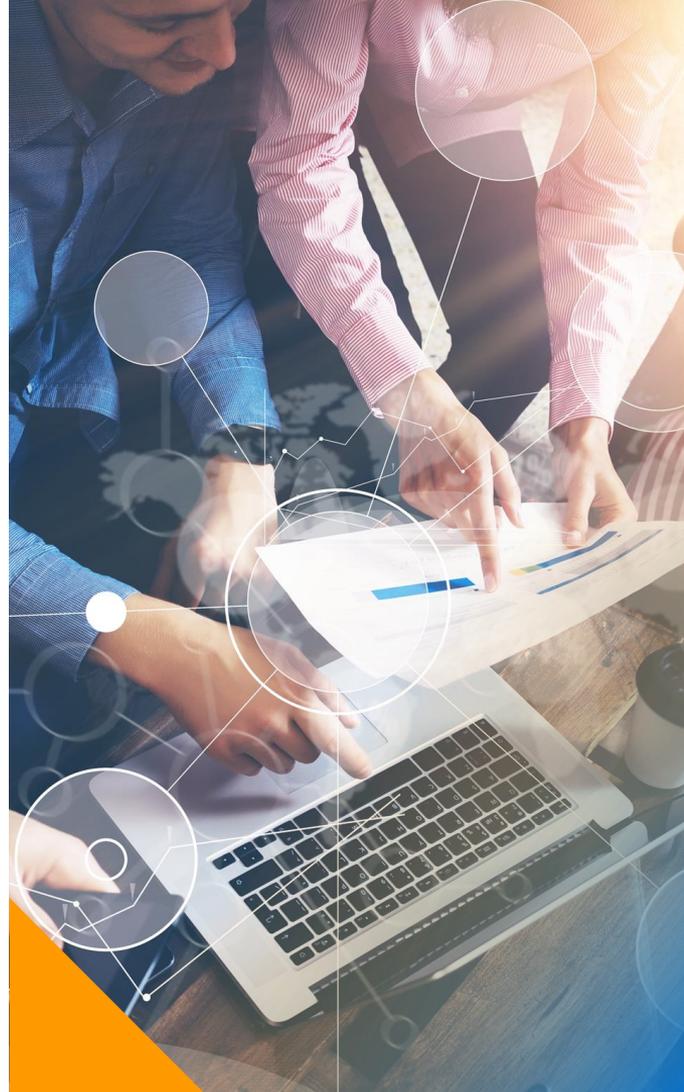




Deep Dive into Survivorship Strategies and Introduction to Fallback Strategy

January 2022

Tatyana Selezneva - QA Engineer
Yakov Goyhman - Engineering Manager
Dmitry Blinov - Sr. Product Manager





- ❑ This session continues a series of Webinars to deep dive the topic of Entity Resolution and Survivorship rules
- ❑ The goal of this session is to
 - ❑ Do a technical deep dive and a demo of each of survivorship strategies
 - ❑ Introduce a concept of the Fallback strategy
 - ❑ Do a demo of the real business use case with use of survivorship and fallback strategies

Agenda

- ★ Introduction and goals of this webinar
- ★ Reltio MDM Entity Resolution
- ★ Technical review of “sourcesForOv” parameter
- ★ Introduction into Fallback strategies
- ★ Deep-dive into Survivorship strategies
 - LUD(Last Update Date) - Demo
 - SRC_SYS - Demo
 - Frequency - Demo
 - Aggregation - Demo
 - OldestValue - Demo
 - MinValue/MaxValue - Demo
 - WinnerEntityCrosswalk - Demo
 - OtherAttributeWinnerCrosswalk - Demo
- ★ Demo of a complex use case

Entity Resolution is in the core of the Multidomain MDM

CUSTOMER BENEFITS



Single source
of truth
at scale



Real-time,
analytics, data
science and UX
for all workloads



Secure &
governed
access via UI,
APIs, streaming
and connectors



SaaS with
portability across
multiple clouds



Zero downtime
maintenance and
disruption-free
upgrades

— The Reltio multi-cloud SaaS Platform —

Data API & UX for Digital Customer & Product Experience

Augmented MDM for Entity 360

Multidomain MDM

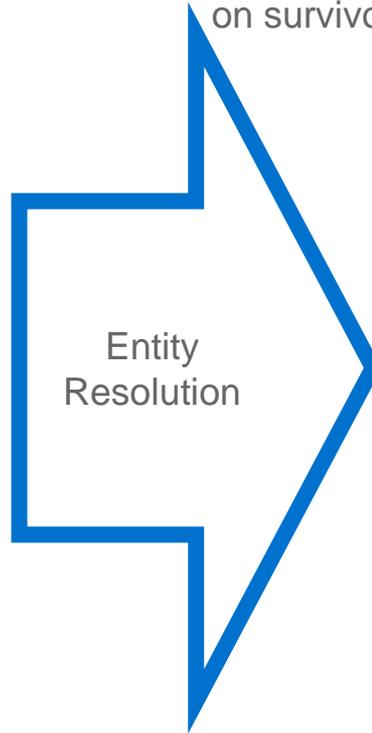
Entity Resolution

Entity Resolution

Entity Resolution happens on the attribute level where each attribute value loses or wins (survive). Survived value is called operational value OV, it is calculated by OV calculator based on survivorship rules set for each attribute type.

Crosswalk(data source) A	
Attribute	Value
First Name	Michael
Last Name	Frasca

Crosswalk(data source) B	
Attribute	Value
First name	Mike
Last Name	F.



Resolved Object		
Attribute	Operational Value	Non Operational Value
First Name	Mike	Michael
Last Name	Frasca	F.

Review of “sourcesForOv” parameter

“sourcesForOv” is an optional field, which contains a list of sources that can take part in OV calculation. The sources which aren't included in this list will be ignored.

“sourcesForOv” can be specified on the group level - in this case sources will be restricted for each mapping inside this group, and on the mapping level - in this case sources will be restricted only for the current mapping.

If the field is specified on both levels - “sourcesForOv” from the mapping level is used.

```
"survivorshipGroups": [  
  {  
    "uri":  
      "configuration/entityTypes/HCP/survivorshipGroups/default",  
    "sourcesForOv": [  
      "configuration/sources/FB",  
      "configuration/sources/TWITTER"  
    ],  
    "default": true,  
    "mapping": [  
      {  
        "attribute":  
          "configuration/entityTypes/HCP/attributes/LastName",  
        "survivorshipStrategy": "OldestValue",  
        "sourcesForOv": [  
          "configuration/sources/AHA"  
        ]  
      }  
    ],  
  },  
]
```

Introduction into fallback strategies

Fallback strategy describes a mapping that will be used over the results of a main or initial strategy in case the main strategy returns a number of winners that does not match the specified condition.

For example, I have an SRC_SYS strategy for the attribute, but it returns too many winners. To shorten them, I can set “LUD” fallback strategy, which will pick up winners from SRC_SYS with the most recent update date.

The condition is determined by “fallbackUsingCriteria” field.

Two types of fallback criteria

MORE_THAN_ONE (default). The fallback will work if a main strategy returns 2 or more winner values. The specified fallback strategy will use ONLY winner values and ONLY winner crosswalks from the PREVIOUS step.

ZERO_OR_MORE_THAN_ONE. If a main strategy returns 0, 2 or more winner values, the logic for determining values used by the fallback strategy is based on the retrieved number of winners:

- in case of 0: all the values from the previous step are regarded by the fallback strategy;
- in case of >1: ONLY winner values and ONLY winner crosswalks from PREVIOUS steps are taken into account.

Sample fallback strategy configuration

```
{
  "mapping": [
    {
      "attribute": "configuration/entityTypes/Individual/attributes/FirstName",
      "survivorshipStrategy": "SRC_SYS",
      //step 1: Check the main strategy. According to the ZERO_OR_MORE_THAN_ONE fallback criteria,
      //If there are zero or more than one winner, then all the values (if zero values won) or only the winner values (if there are more than one
winner) are passed to the fallback section.
      "fallbackUsingCriteria": "ZERO_OR_MORE_THAN_ONE",
      "fallbackStrategies": [
        {
          "attribute": "configuration/entityTypes/Individual/attributes/FirstName",
          "survivorshipStrategy": "LUD",
          //step 2: The strategy used depends on the number of winners from the previous step. If the LUD strategy returns more than one winner,
the next fallback strategy will take part in the Operational Value (OV) calculation.
          "fallbackUsingCriteria": "MORE_THAN_ONE",
          "fallbackStrategies": [
            {
              "attribute": "configuration/entityTypes/Individual/attributes/FirstName",
              "survivorshipStrategy": "MinValue" //this step is required only if the LUD strategy returns more than one winner value (in
accordance with the MORE_THAN_ONE fallback criteria).
            },
            ...
          ]
        }
      ]
    }
  ]
}
```

Types of survivorship strategies

Data Source based

- LUD
- SRC_SYS
- OldestValue
- OtherAttributeWinnerCrosswalk
- WinnerEntityCrosswalk

Data or Value based

- Frequency
- Aggregation
- MinValue
- MaxValue

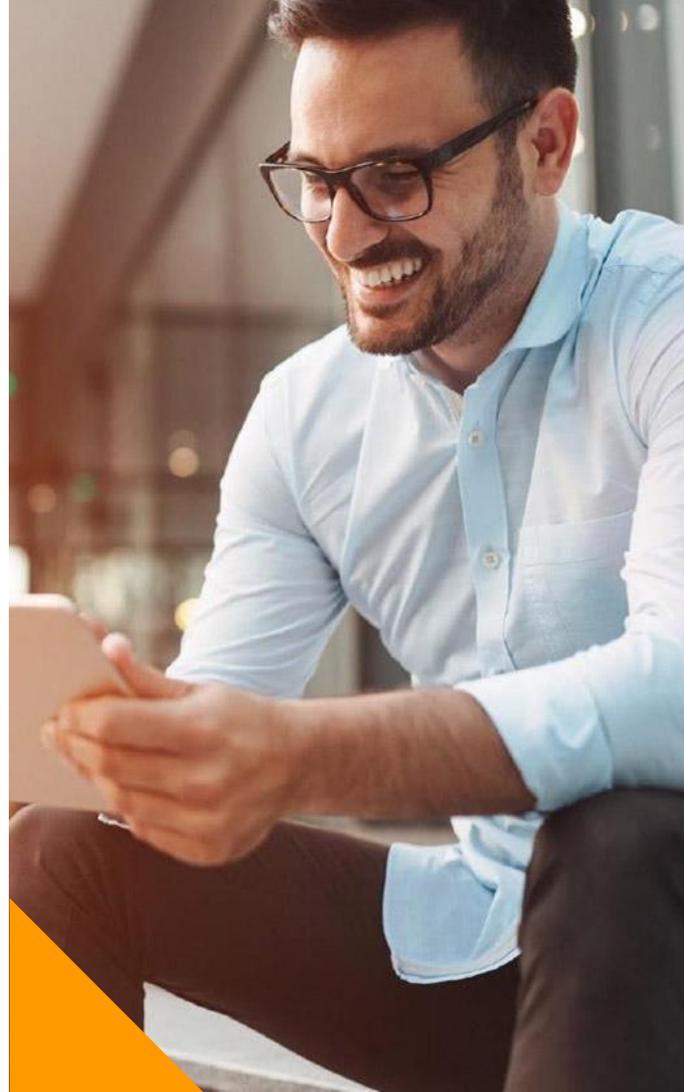
The screenshot shows a data management interface with a table and a dropdown menu. The table has columns for 'First Name', 'Source System', and 'Value'. The first row contains 'Thomas', 'Facebook', and 'Thomas'. A dropdown menu is open over the 'Source System' column, showing a list of survivorship strategies. The 'Recency' option is selected and highlighted.

ATTRIBUTES	APPLIED OPERATIONAL VALUES	WINNER SOURCE SYSTEM	RULE TYPE	COUNT	ATTRIBUTE VALUES
+ Add					
First Name:	Thomas	Facebook	Recency	1	Thomas

- ✓ Recency
- Reltio Cleanser or Nothing
- Recency
- Frequency
- Aggregation
- Source system
- Oldest value
- Minimum value
- Maximum value
- Other Attribute Winner Crosswalk
- WinnerEntityCrosswalk



Data Source based or Crosswalk based strategies



“Make all values with the most recent update date to be winners for this attribute”

LUD - Last Update Date

1. It takes all the crosswalks bound to each value of the regarded attribute;
2. Then it calculates the maximum value of “updateDate”, “singleAttributeUpdateDates” and “reltioLoadDate” for each crosswalk.
3. Then it finds the maximum value among all the values from Step #2. A corresponding crosswalk becomes a winner crosswalk and all the values, provided by this crosswalk - winner values.

This strategy is a default one. It means that if you don't specify any mapping for an attribute, it will use LUD strategy to calculate OV values for this attribute.

Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F0A72Ymv_ps~profile

Entity: 0A72Ymv

Attribute: FirstNameLUD

Post 1: Mike (hcp_lud_A) + Michael (hcp_lud_B)

Update: Michael -> MICHAEL

Michael wins

SRC_SYS

“Make values, belonging to a source with the highest priority to be winners”

This strategy is based on “sourcesUriOrder” field.

Primarily, “sourcesUriOrder” is searched at the mapping level. If it doesn't have such an element, we search for it on the group level.

After that, we take the source with the highest priority from this list and search for values of the attribute that belong to this source. If they are found - we consider them as winners, if not - we continue looking through the list until we find such sources or the list is over. In case the attribute doesn't have the sources specified in “sourcesUriOrder” list from the mapping or group level, we make up a list of sources from the “priority” fields of the “sources” section, and then check it the same way, as we checked the list of sources taken from the mapping or group level.

There can be a situation when there are no winners at all, but it depends on if the advanced OR not advanced behavior is used and if there are fallback strategies set up for the mapping.

Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F08jrJjc_ps~profile

Entity: 08jrJjc

Attribute: FirstNameSRCSYS

```
Config: "sourcesUriOrder": [  
    "configuration/sources/IREP",  
    "configuration/sources/HCOS",  
    "configuration/sources/HMS",  
    "configuration/sources/DEA"  
]
```

Mike_FB

Mike_HMS => wins

Mike_DEA

“Make values that came first to be winners”

Oldest Value

1. It takes all the crosswalks bound to each value of the regarded attribute;
2. Then it calculates the minimum value of “createDate” for all the crosswalks from Step #1 and takes all the values with the oldest create date as winners.

Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13he~entities%2F0A72hJR_ps~profile

Entity: 0A72hJR

Attribute: FirstNameOldestValue

Post 1: Mike (hcp_lud_A) + Michael (hcp_lud_B)

Update: Michael -> MICHAEL

Mike wins

Other Attribute Winner Crosswalk

“Make attributes from the source that wins in the other attribute to be winners for this attribute”

It takes winner sources for the attribute from another. It is useful when one attribute depends on another.

The strategy has a mandatory field "primaryAttributeUri" to specify a URI of the "main" (or "primary") attribute. Primary Attribute OV calculates before the dependent one in accordance with the strategy, specified for this attribute.

E.g. "LastName" has "OtherAttributeWinnerCrosswalk" strategy and depends on "FirstName".

"FirstName" has "SRC_SYS" strategy in its mapping.

Primarily, "FirstName" is calculated by "SRC_SYS" strategy. After the winner sources are defined for "FirstName", the values with these sources become winners in "LastName".

If a primary attribute doesn't have winner crosswalks, the dependent attribute also doesn't have winner (OV) values.

It's possible to have primary attribute which survivorship mapping also OtherAttributeWinnerCrosswalk. There is only a requirement to not create cyclic dependencies: attribute can't depend on itself directly (using current attribute as primaryAttributeUri) or indirectly (through chain of other attributes by primaryAttributeUri).

Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13he~entities%2FOA72ppx_ps~profile

Entity: 0A72ppx

WinnerEntityCrosswalk

“Make all values of the attribute from the winner entity to be winners”

- Before a merge operation, all the value of the attribute has OV=true.
- If a winner is set explicitly in a merge operation, then it's values will be OV=true.
- If a winner is NOT set explicitly in a merge operation, the oldest entity will be chosen as a winner and its values will be OV=true

Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13he~entities%2F5u0G0OE_ps~profile

Entity: 5u0G0OE

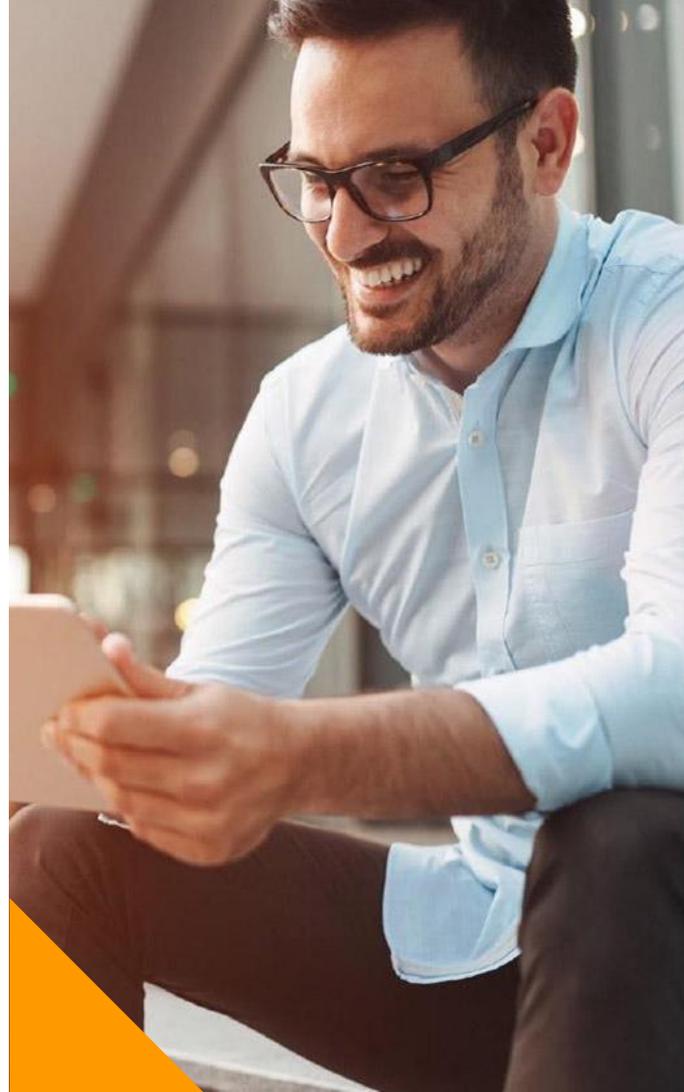
Attribute: FirstNameWinnerEntityCrosswalk

E1: Mike (from Entity1)

E2: Michael (from Entity2), E2 wins => this value wins



Data based or Value Based strategies



Frequency

“Make values that came from the most number of different sources to be winners”

- If more than one value win, LUD strategy is used to find final winners.
- If this strategy is used for a Nested or Reference attribute:
 - As we consider Nested and Reference attributes as folders, or containers, we need to set one sub-attribute, which can unambiguously define our Nested/Reference attribute (E.g. for “Identifiers” nested attribute is ID). A URI of such sub-attribute has to be placed in **comparisonAttributeUri** field. The “Frequency” strategy will use the specified attribute to define winners for the whole Nested/Reference attribute.
 - **comparisonAttributeUri** is obligatory for Nested and Reference attributes. Otherwise, a validation error is shown.
 - If a value of the attribute, specified in **comparisonAttributeUri**, doesn't exist, the value of the nested attribute is considered as NULL.

Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13he~entities%2F08jrame_ps~profile

Entity: 08jrame

Attribute: FirstNameFrequency

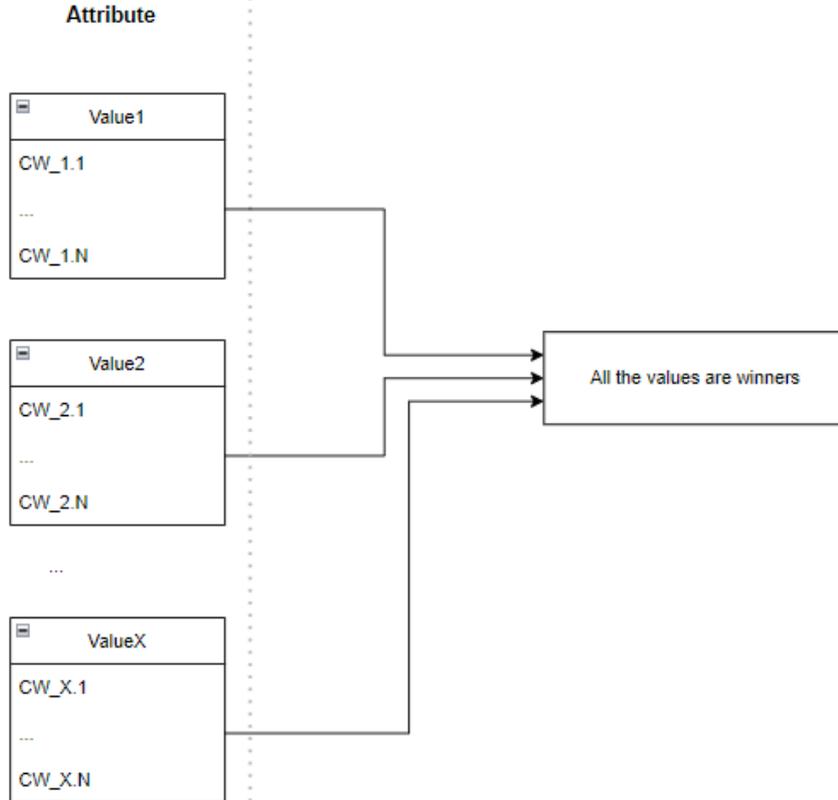
Mike (FB) => win

Mike (HMS) => win

MIKE (NPI)

“Make all values to be winners”

Aggregation



Demo scenario

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13he~entities%2F5u0G4eU_ps~profile

Entity: 5u0G4eU

Attribute: FirstNameAggregation

Mike (FB)

Mike (HMS)

MIKE (NPI)

Win: Mike x1, MIKEx1

“Make min/max values to be winners”

MinValue/MaxValue

Calculation of minimum value is based on attribute type:

- Numeric - the minimum value is the smaller numeric value
- Date - the minimum value is the minimum date value
- Boolean - true > false
- String - the minimum value is based on the lexicographical sort order of the strings

If you need to sort values in a way that differs from the default sorting, you can use “sortAs” field, implemented specially for this strategy. It contains the type of object, which a value of the attribute will be cast to.

*As for the “Frequency” strategy, “MinValue” requires **comparisonAttributeUri** in case of Nested and Reference attributes, otherwise, a validation error is shown. If a value of the attribute, specified in **comparisonAttributeUri**, doesn't exist, the value of the nested attribute is considered as NULL. It's always counted as the maximum compared with any other value.*

Demo for min value

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F5u0GHRG_ps~profile

Entity: 5u0GHRG

Attribute: FirstNameMinValue

Bob => wins

Bobby

Bob J.

Demo for max value

https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F5u0GUE2_ps~profile

Entity: 5u0GUE2

Attribute: FirstNameMaxValue

Bob

Bobby => wins

Bob J

Demo

Scenario:

- Addresses as a nested attribute calculated by Max strategy Max
- Check sub-attribute AddressRank (using comparisonAttributeURI), as it has number values but the attribute type is - String
- in the configuration add sortAs: Int

- POST 1:
- 548 Perry St, AddressRank = 10, B=> wins on 1st step
- 81 Filmore St, AddressRank = 1,HMS
- 6542 Logan Square, AddressRank = 10, NPI => no match due to sourcesForOv
- 700 Church Rd, AddressRank = 7, HMS

- POST 2:
- Add 466 Main St, AddressRank = 10, HMS.
- So now we have 2 winners and fallbackUsingCriteria not set (=MORE_THAN_ONE), fallback strategy kicks in by SRC_SYS
- Winner sources are - FB, HMS - not matching `sourcesUriOrder`, by ZERO_OR_MORE_THAN_ONE we then move to the next fallback - LUD.
- On this stage 466 Main St, wins as it was added last

- POST3
- Update the very first address (548 Perry St, AddressRank = 10) using the cumulative update
- singleAttributeUpdateDate changed and it wins

Demo

```
{  
  "attribute": "configuration/entityTypes/HCP/attributes/Addresses",  
  "survivorshipStrategy": "MaxValue",  
  "sortAs": "Int",  
  "comparisonAttributeUri": "configuration/entityTypes/HCP/attributes/Addresses/attributes/AddressRank",  
  "sourcesForOv": [  
    "configuration/sources/AHA",  
    "configuration/sources/FB",  
    "configuration/sources/HMS",  
    "configuration/sources/DEA"  
  ],  
  "fallbackStrategies": [  
    {  
      "attribute": "configuration/entityTypes/HCP/attributes/Addresses",  
      "survivorshipStrategy": "SRC_SYS",  
      "sourcesUriOrder": [  
        "configuration/sources/AHA",  
        "configuration/sources/DEA"  
      ],  
      "fallbackUsingCriteria": "ZERO_OR_MORE_THAN_ONE",  
      "fallbackStrategies": [  
        {  
          "attribute": "configuration/entityTypes/HCP/attributes/Addresses",  
          "survivorshipStrategy": "LUD"  
        }  
      ]  
    }  
  ]  
}
```

Link to demo:

[https://eutst-](https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F08jqpto_ps~profile)

[01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F08jqpto_ps~profile](https://eutst-01.reltio.com/ui/SeleznevaTatyana/#p~com.reltio.plugins.entity.perspective.HCP_w~13h_e~entities%2F08jqpto_ps~profile)

Entity ID: 08jqpto